Generative Planning

for Temporally Coordinated Exploration in Reinforcement Learning

Haichao Zhang, Wei Xu and Haonan Yu

Horizon Robotics

What is **Generative Planning?**

- a perspective on extending standard model-free RL algorithms for intentional exploration;

What is **Generative Planning?**

- a perspective on extending standard model-free RL algorithms for intentional exploration;

- achieved by leveraging a connection between *planning* and *temporally extended exploration*, generating planned actions not only for the current step, but also for several future steps.

Why Generative **Planning?**

Why Generative **Planning?**

Intentional Exploration

- standard model-free RL explores inefficiently at a single time-step level
- planning is better at reasoning over long horizons and can be used for intentional exploration

Why Generative **Planning?**

Intentional Exploration

- standard model-free RL explores inefficiently at a single time-step level
- planning is better at reasoning over long horizons and can be used for intentional exploration

Interpretation

- since the plan can be interpreted as the intent of the agent from *now* into the *future*, it offers a more informative and intuitive signal for interpretation

Why Generative **Planning?**

Intentional Exploration

- standard model-free RL explores inefficiently at a single time-step level
- planning is better at reasoning over long horizons and can be used for intentional exploration

Interpretation

- since the plan can be interpreted as the intent of the agent from *now* into the *future*, it offers a more informative and intuitive signal for interpretation

Adaptiveness

- by planning a sequence of actions online, it is potentially more effective than commonly used action repeat strategy, which is a non-adaptive form of plans

Why **Generative** Planning**?**

Why **Generative Planning**?

Standard Planning

- optimizes a sequence of actions by online minimization of a cost function computed together with a model
- computationally demanding and requires access to a model

Why **Generative Planning**?

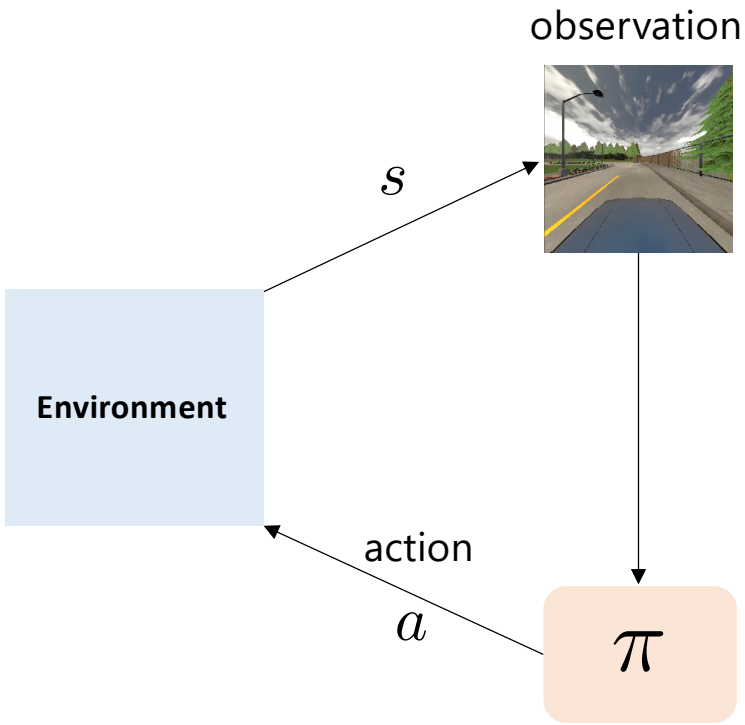
Standard Planning

- optimizes a sequence of actions by online minimization of a cost function computed together with a model
- computationally demanding and requires access to a model

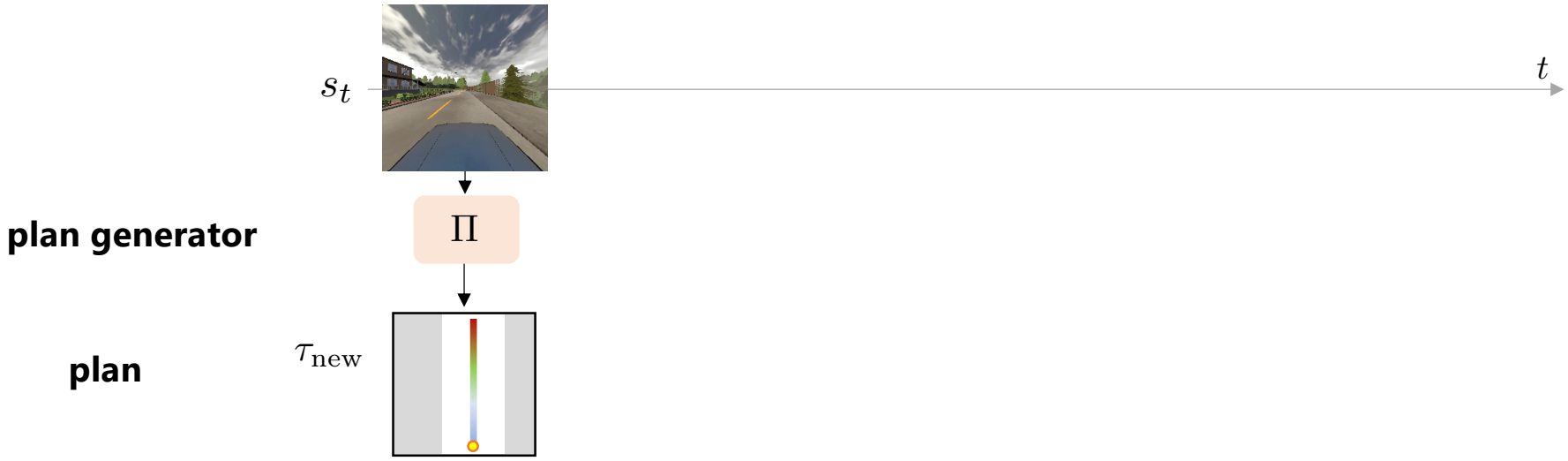
Generative Planning

- uses a generative network to generate plans
- amortizes the expensive online optimization into training, thus is more computational friendly
- does not require an explicit dynamics model

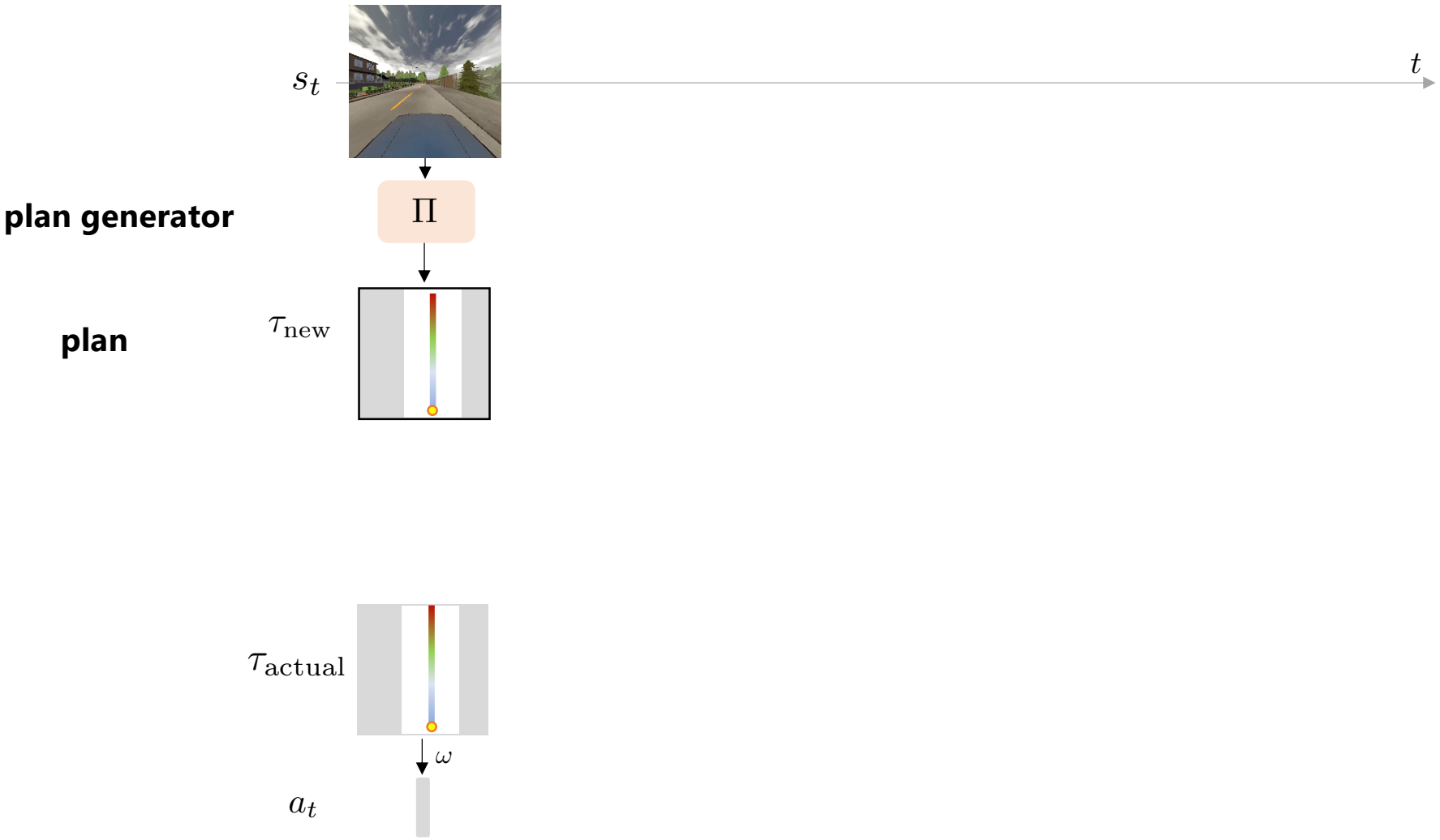
Standard RL



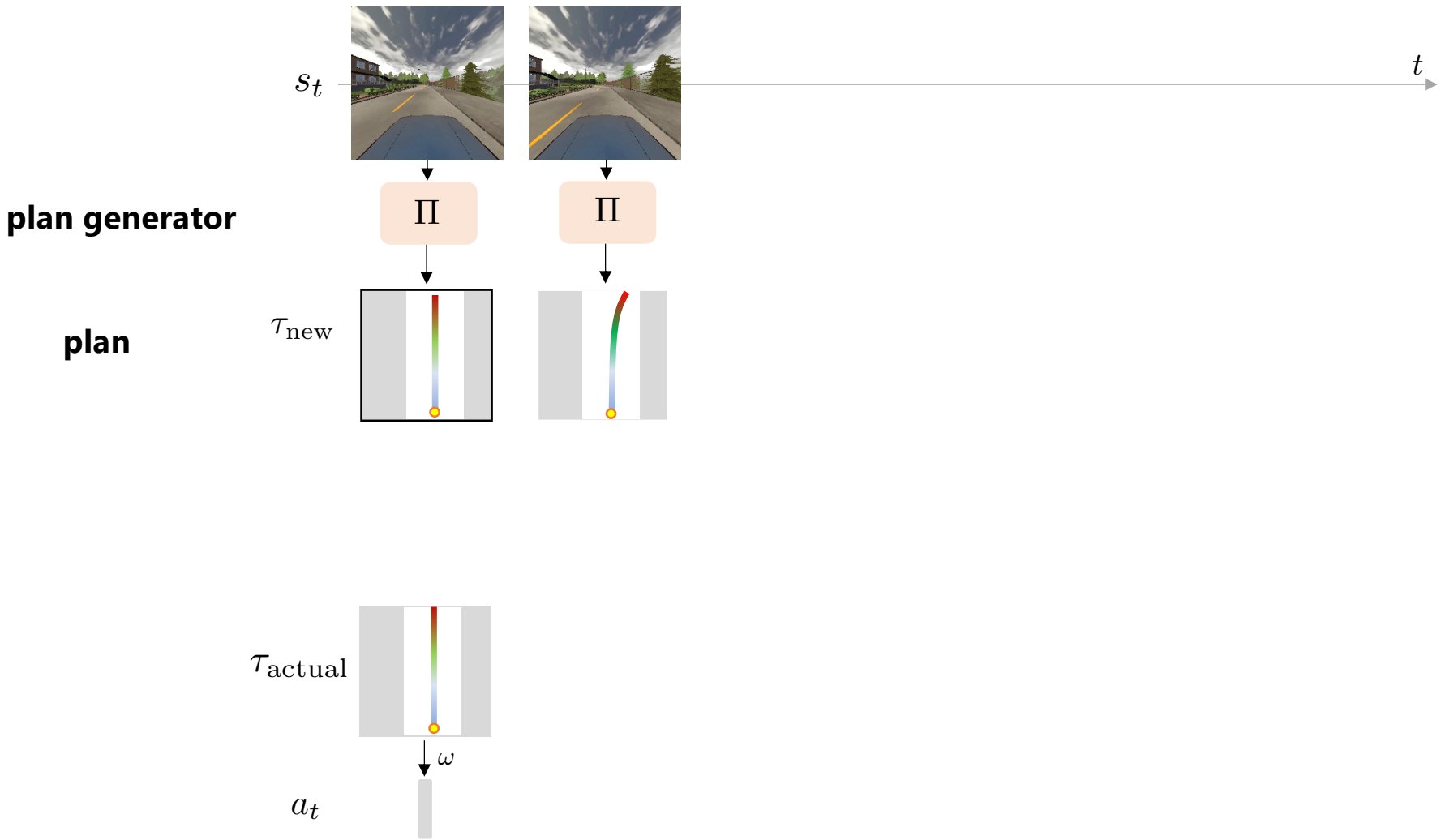
Generative Planning Method (GPM)



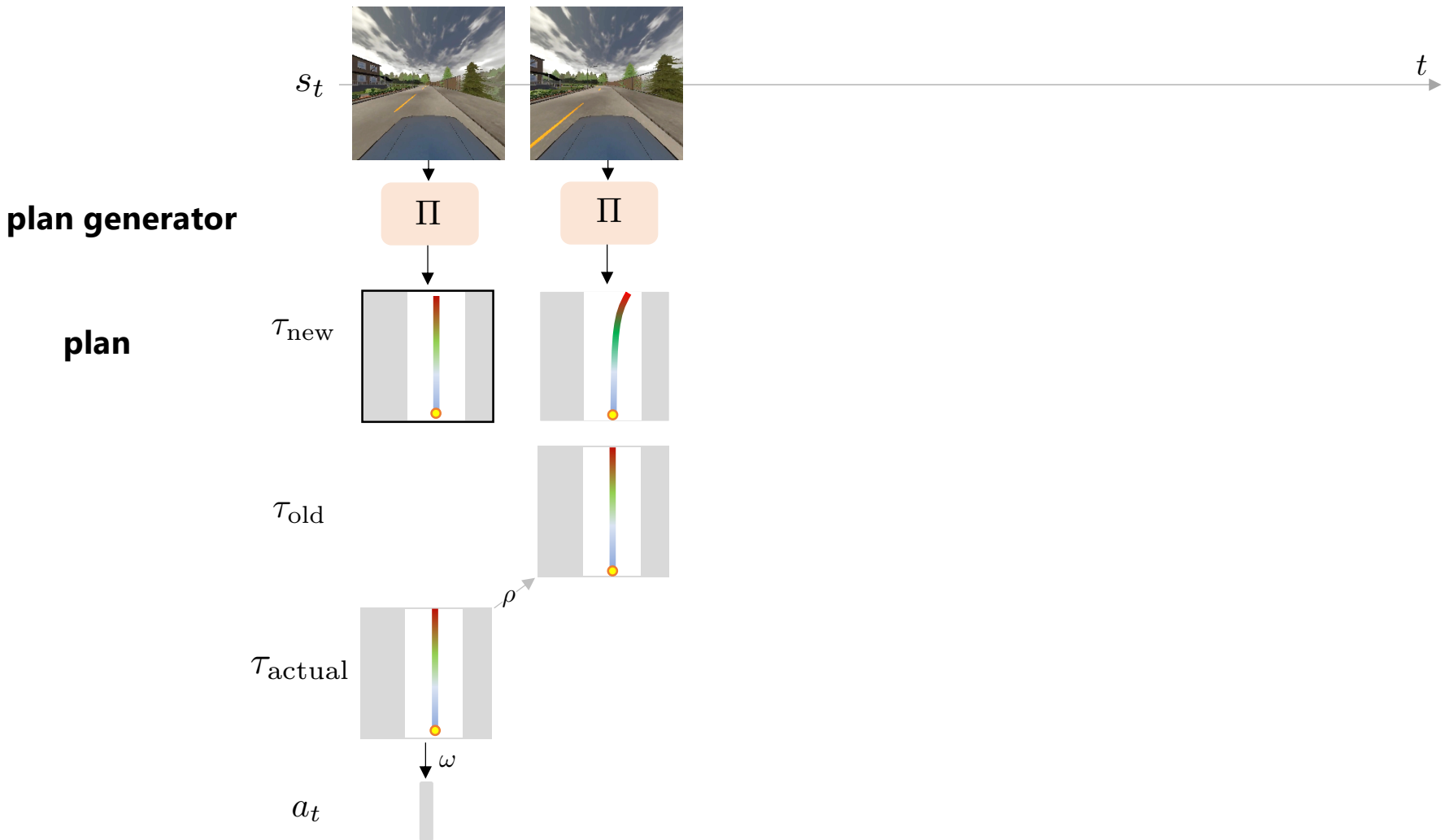
Generative Planning Method (GPM)



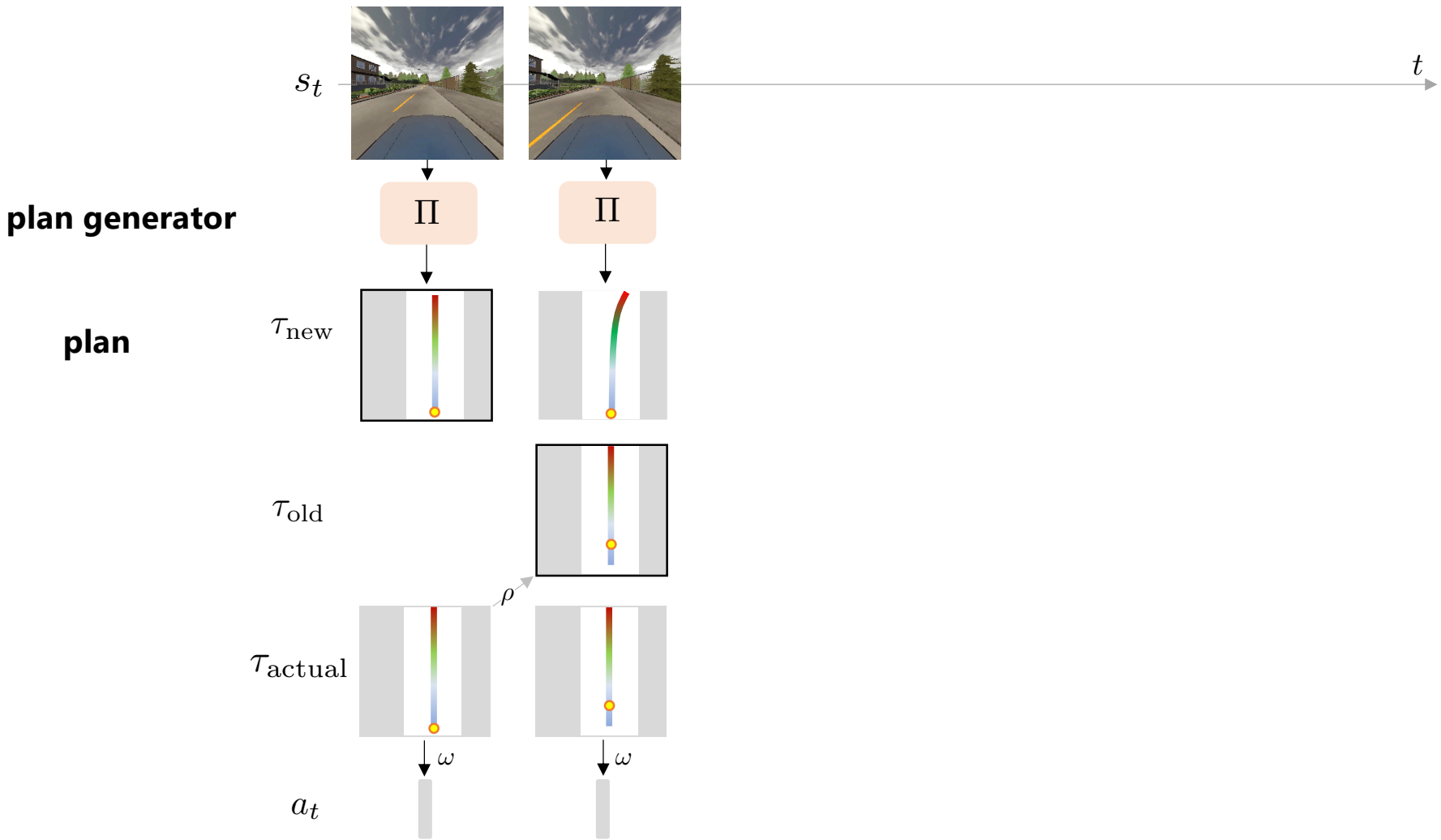
Generative Planning Method (GPM)



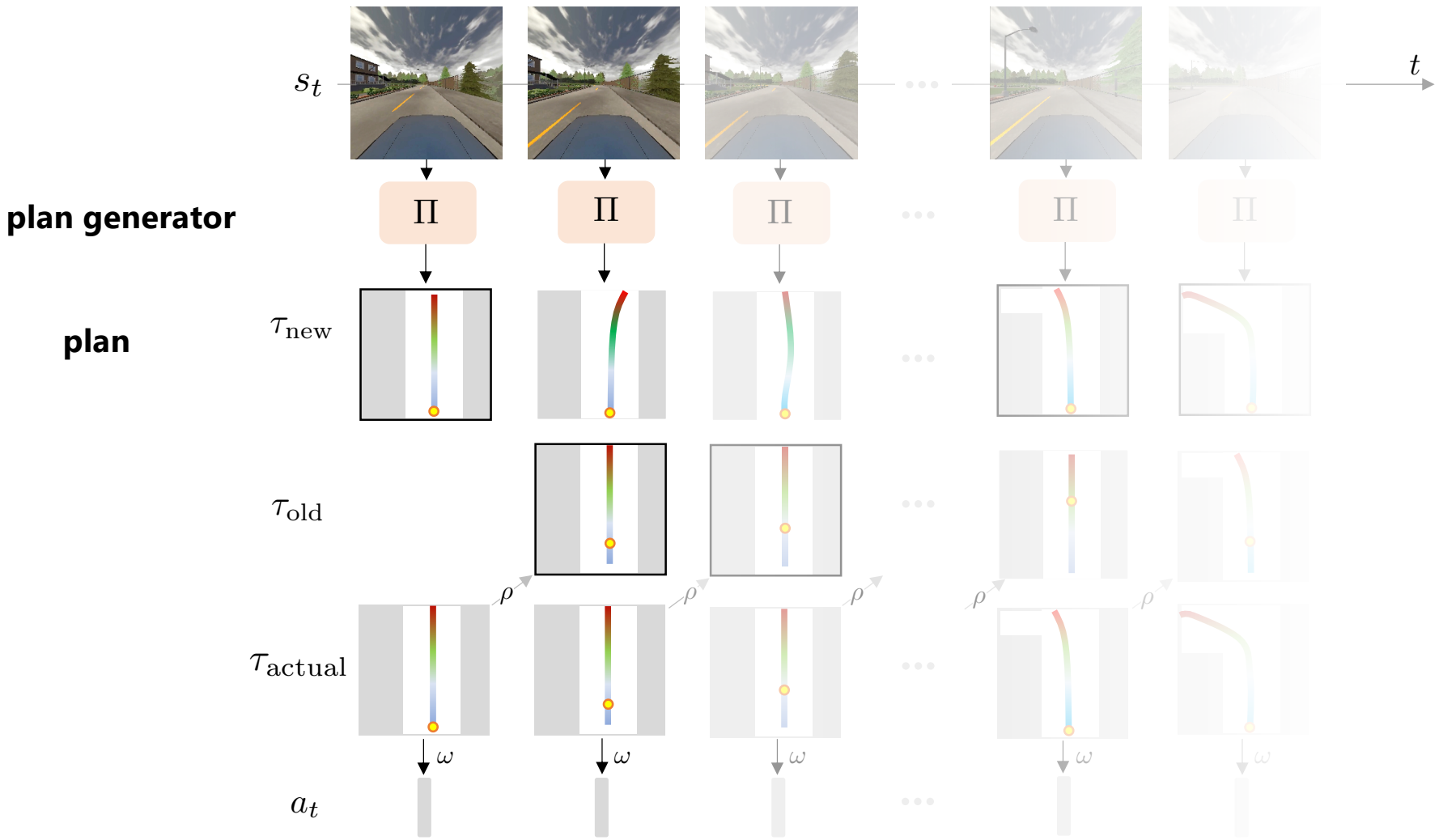
Generative Planning Method (GPM)



Generative Planning Method (GPM)

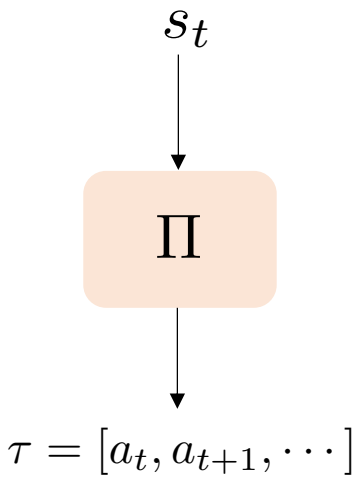


Generative Planning Method (GPM)

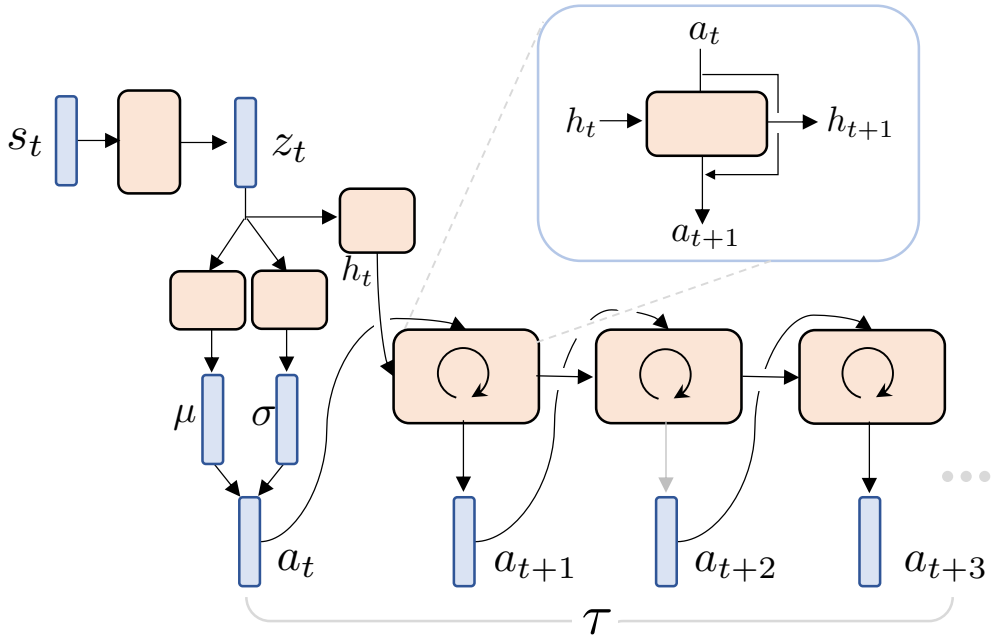


Plan Generator Modeling

Plan Generator (actor)



modeled with a stochastic autoregressive network to capture the inherent temporal relations between consecutive actions



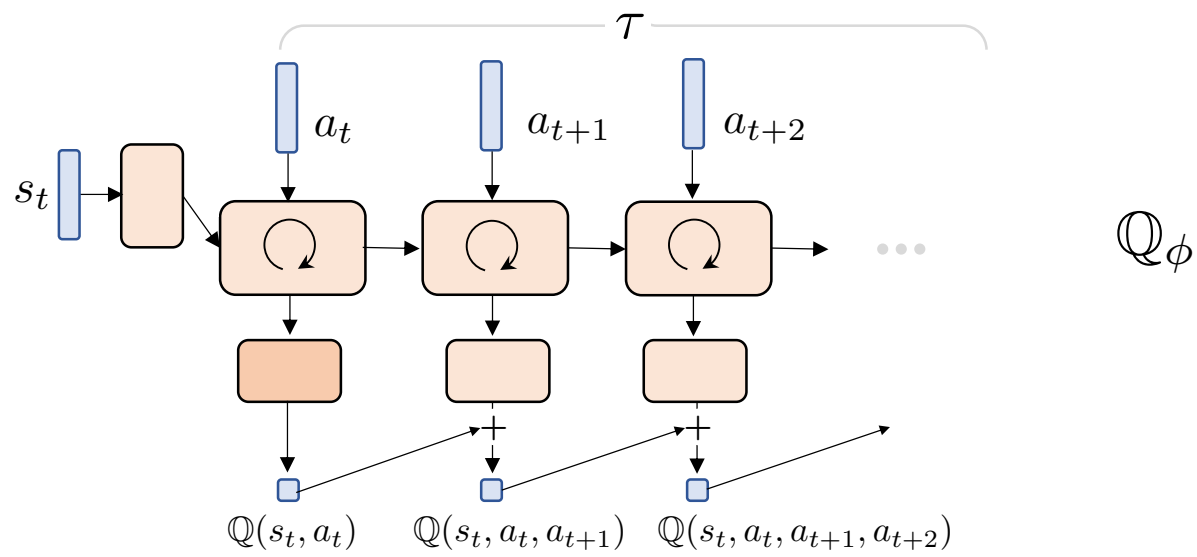
Π_θ

Plan Value Function

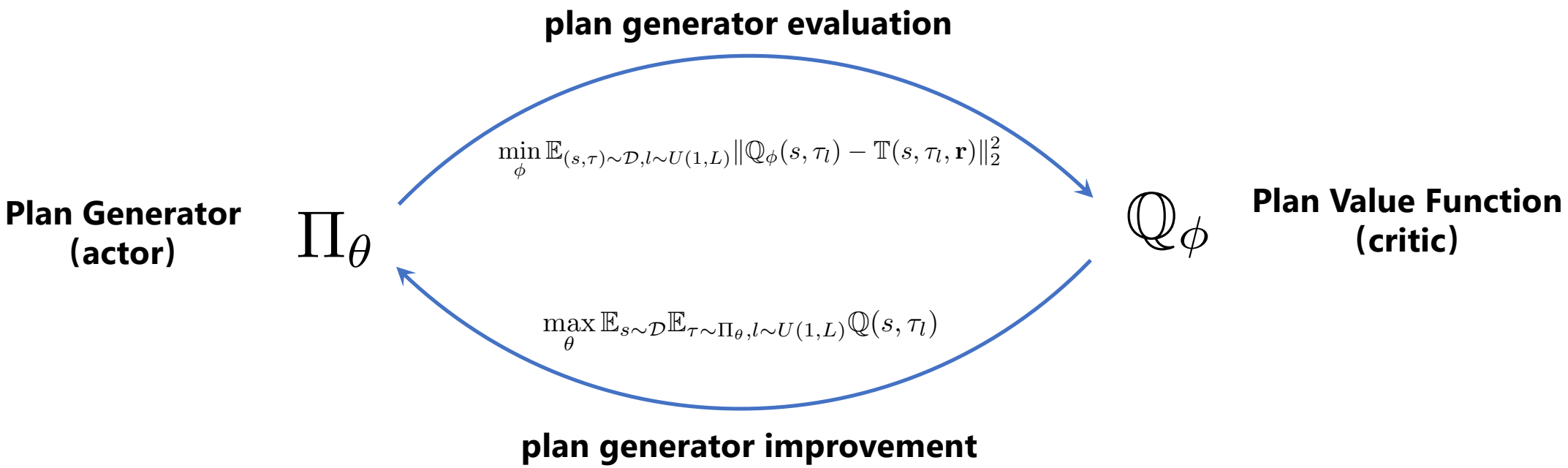
Plan Value Function (critic)

$$\mathbb{Q}(s_t, \tau_l) = r_{t+1} + \gamma r_{t+2} + \cdots + \gamma^l \mathbb{E}_{s' \sim T(s,\tau_l), \tau' \sim \pi_\theta(s')} \left[\mathbb{Q}(s', \tau') \right]$$

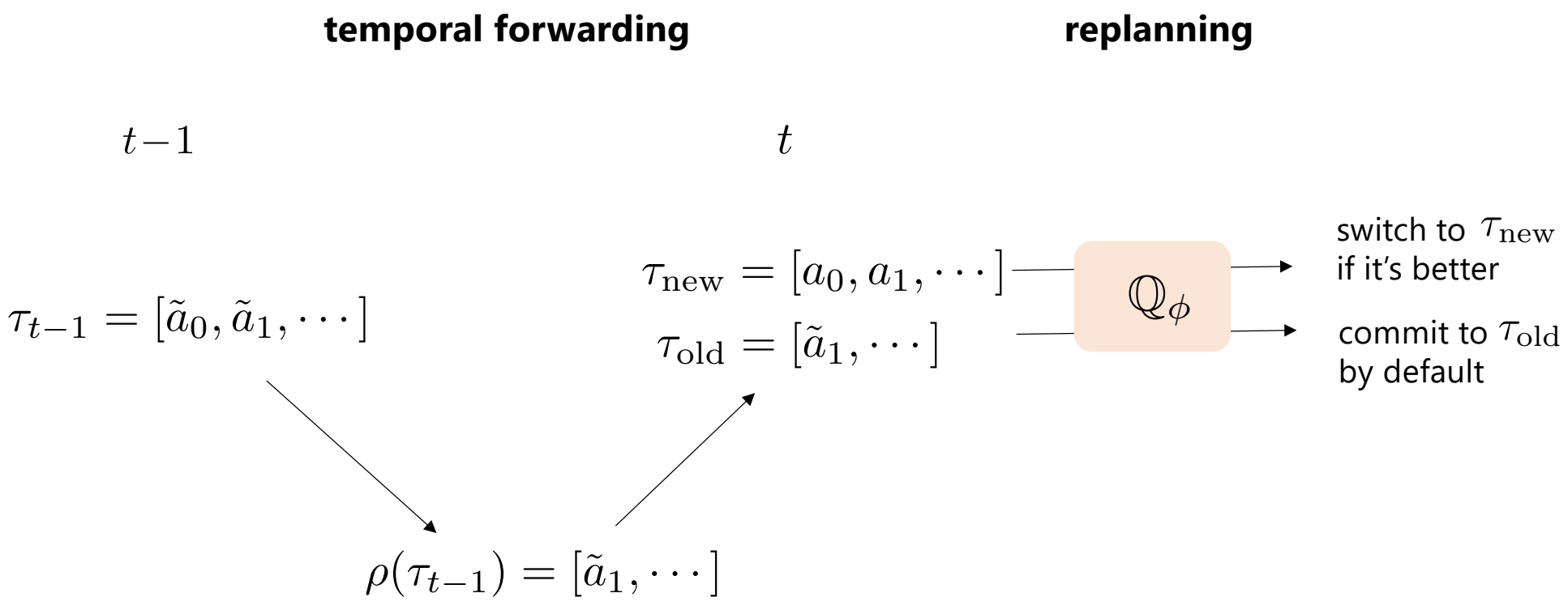
- takes state and plan as input and outputs a sequence of values along the plan
- modeled with RNN
- used in plan generator training and replanning



Plan Value Learning and Generator Training



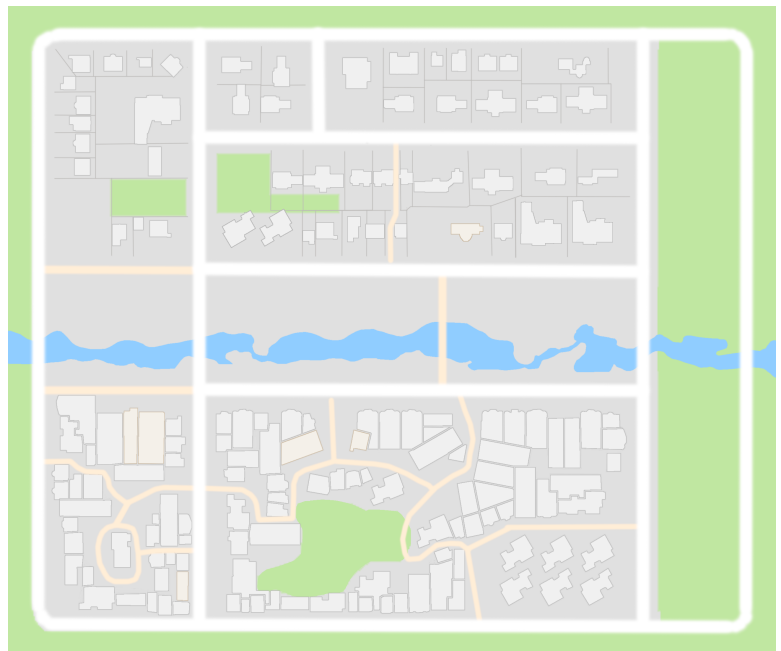
Temporal Forwarding and Replanning



Autonomous Driving Task

- **Simulator:** CARLA*
- **Task**: navigate to the goal position following the route and stay still after arrived at the goal location

Town 01

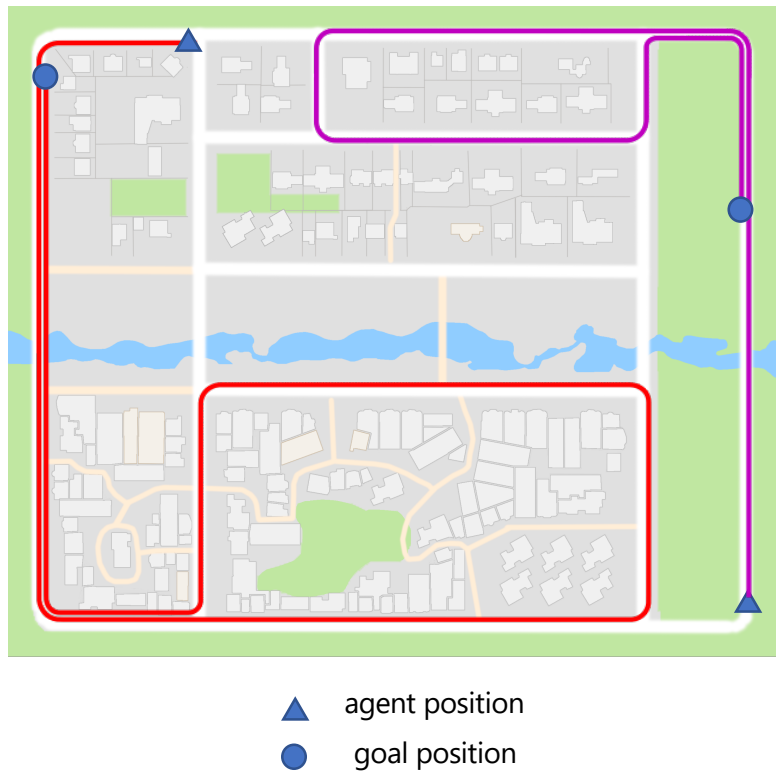


*Dosovitskiy et al., CARLA: An Open Urban Driving Simulator, CoRL 2017

Autonomous Driving Task

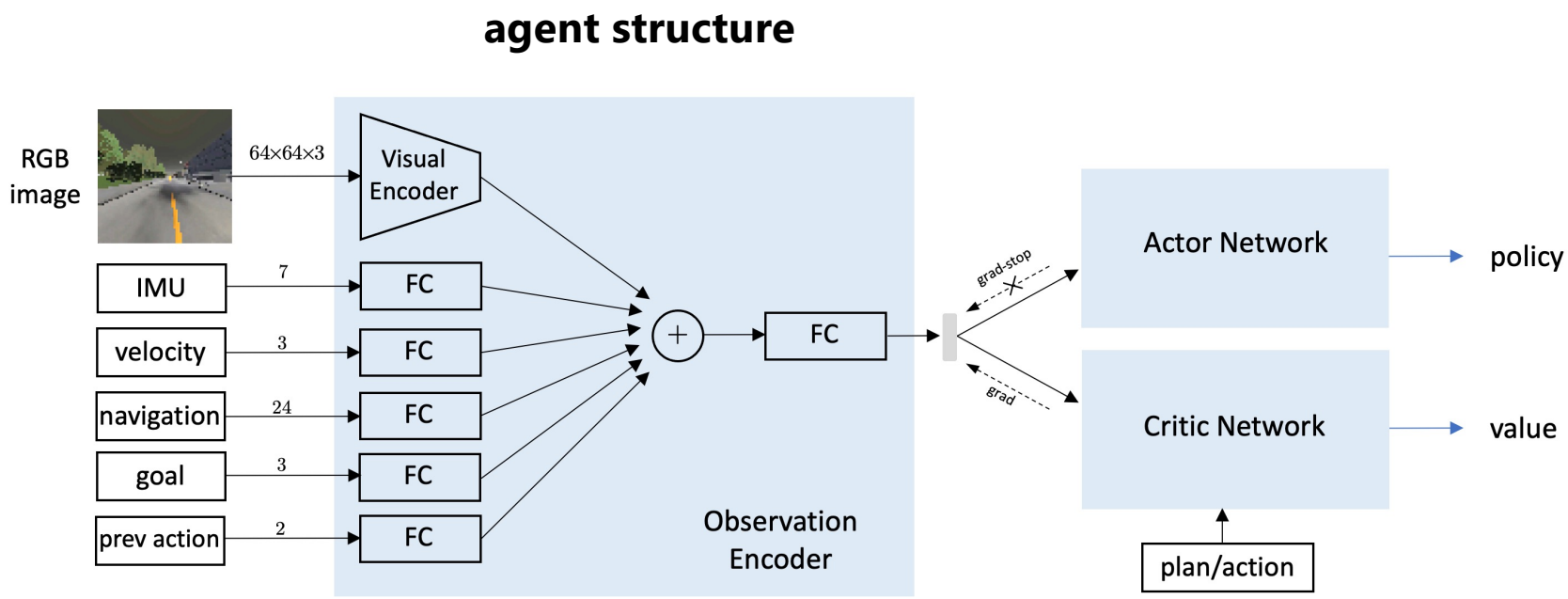
- **Simulator:** CARLA*
- **Task**: navigate to the goal position following the route and stay still after arrived at the goal location
- **Route**: randomly select agent and goal positions from the set of valid waypoints
  - route generated by a route planner given the two points
  - introduces diversity and covers representative driving scenarios between different routes, e.g. different number of turns, directions for turning, static and dynamic objects.

Example Routes

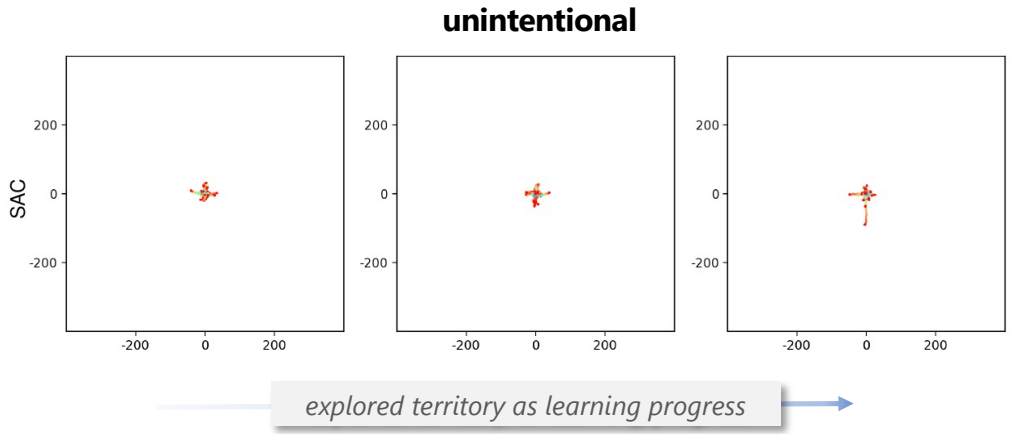


*Dosovitskiy et al., CARLA: An Open Urban Driving Simulator, CoRL 2017

Autonomous Driving Task

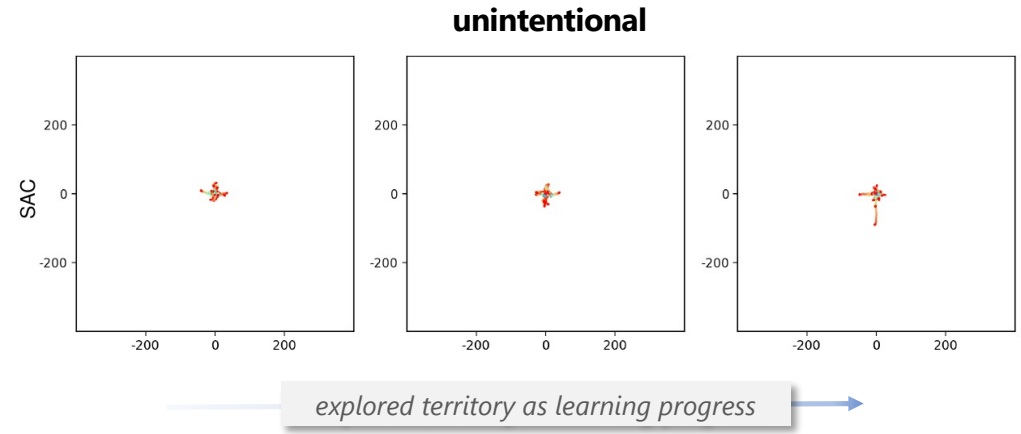


Intentional Exploration

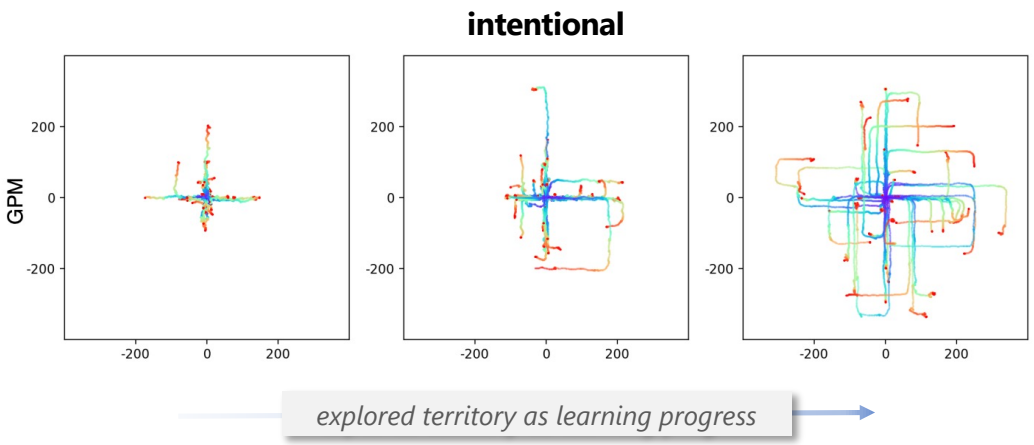


unintentional exploration: explores inefficiently as learning progress

Intentional Exploration



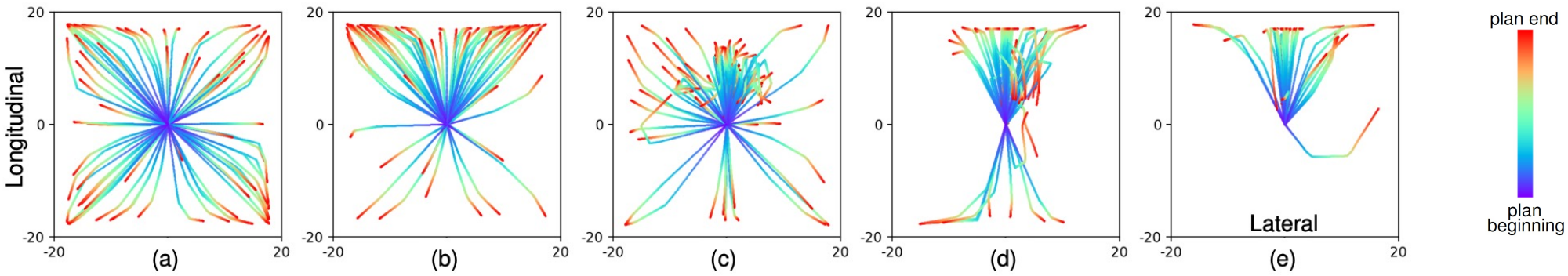
unintentional exploration: explores inefficiently as learning progress



intentional exploration: explores more efficiently and covers a larger territory after the same number of learning steps

Emerged Interpretable Plans

Interpretable behavior: plans emerged during learning are intuitive for interpretation



plans with random directions

plans with more forward directions

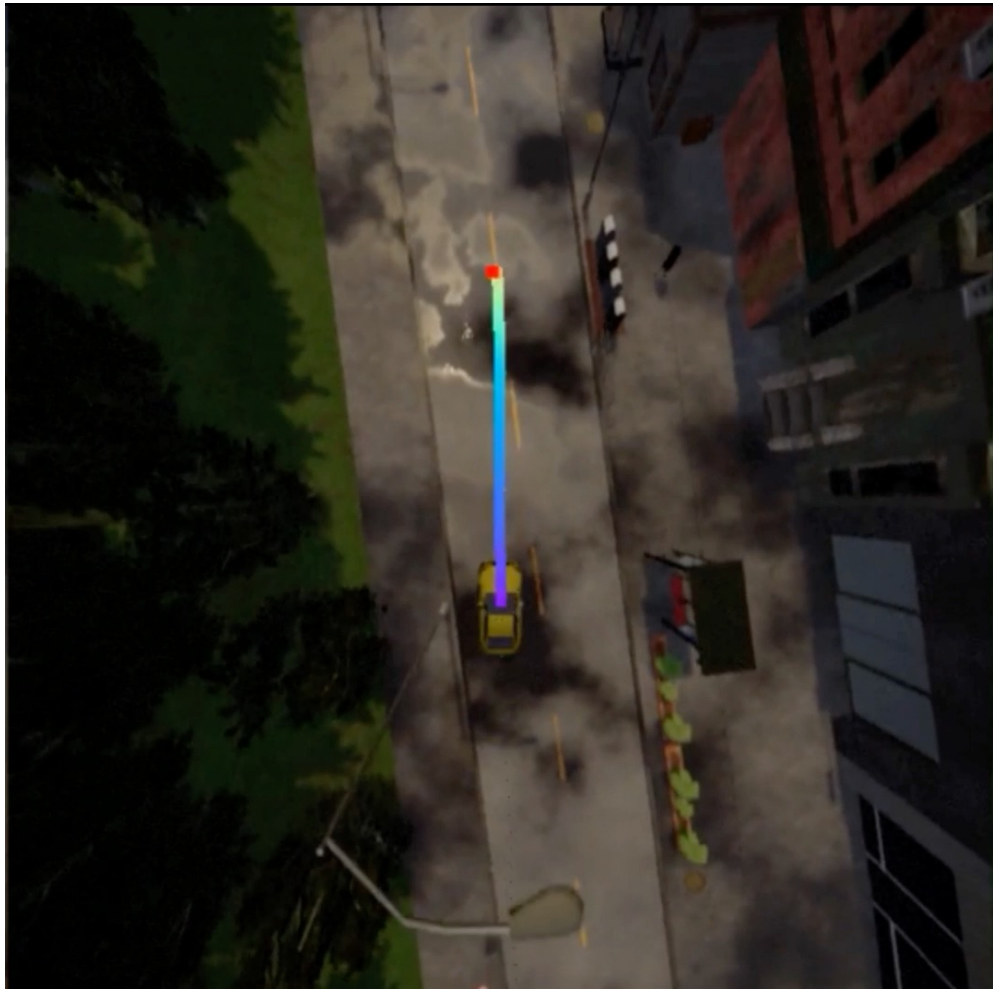
plans with speed adjustment

plans for turning

plans for moving forward and turning

learning progress

Autonomous Driving with GPM



Summary

- **Generative Planning**
  - A perspective on extending standard model-free RL algorithms with intentional exploration;
  - GPM as an instantiation of the idea;
  - Intentional exploration;
  - Interpretable plans.