How to do MCTS on a POMDP?

$Q[(s,a)]$

$N[(s,a)]$

Online POMDP Methods

Approximate POMDP Solutions

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

↓

Offline

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Online

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Previously

Online

Formulation Approximations

(solve a slightly different problem)

Approximate POMDP Solutions

Numerical Approximations

(approximately solve original problem)

Offline

Previously

Online

Formulation Approximations

(solve a slightly different problem)

Last Time

Approximate POMDP Solutions

QMDP
Certainty Equivalence

Numerical Approximations

(approximately solve original problem)

Offline

Previously

SARSOP

Online

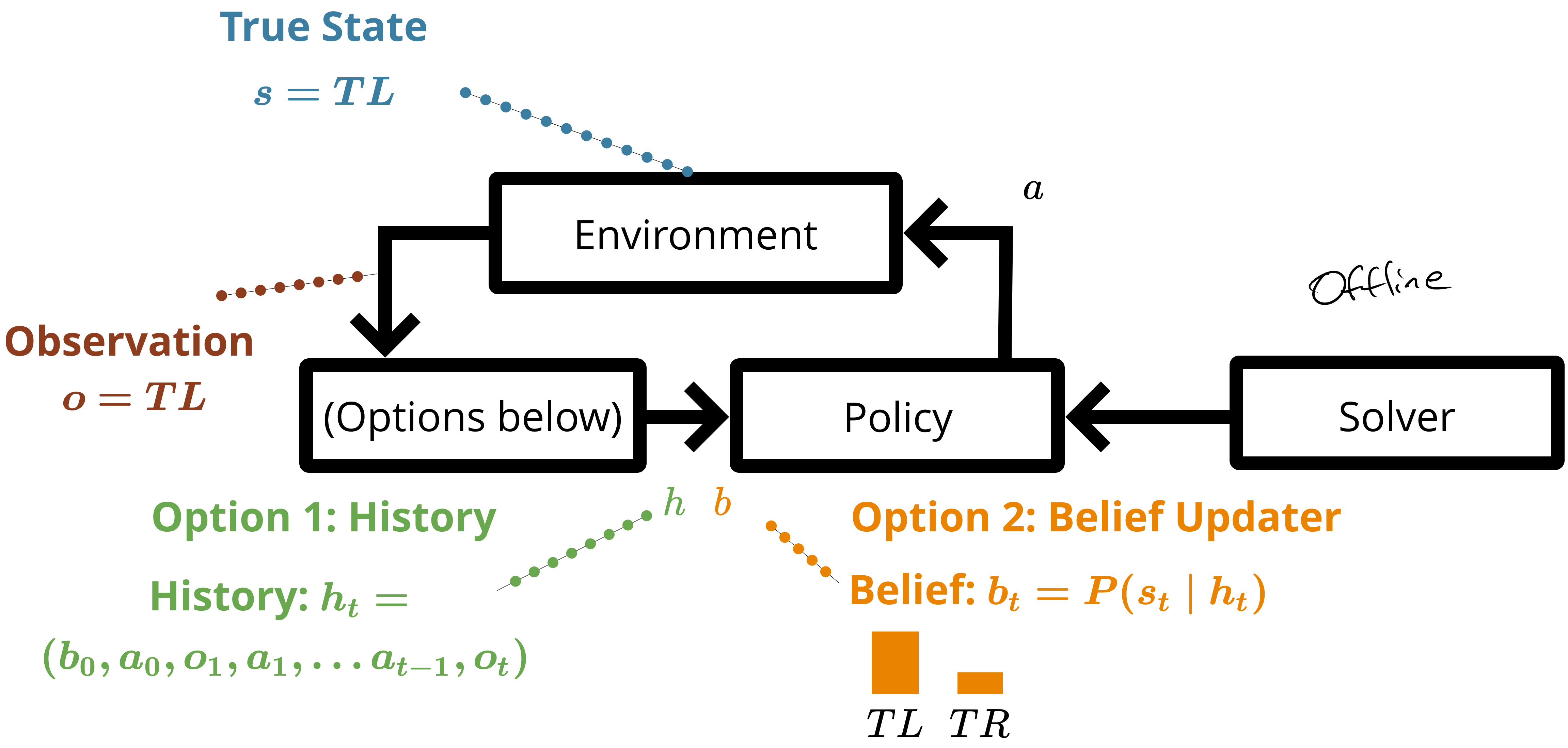
Today!

Formulation Approximations

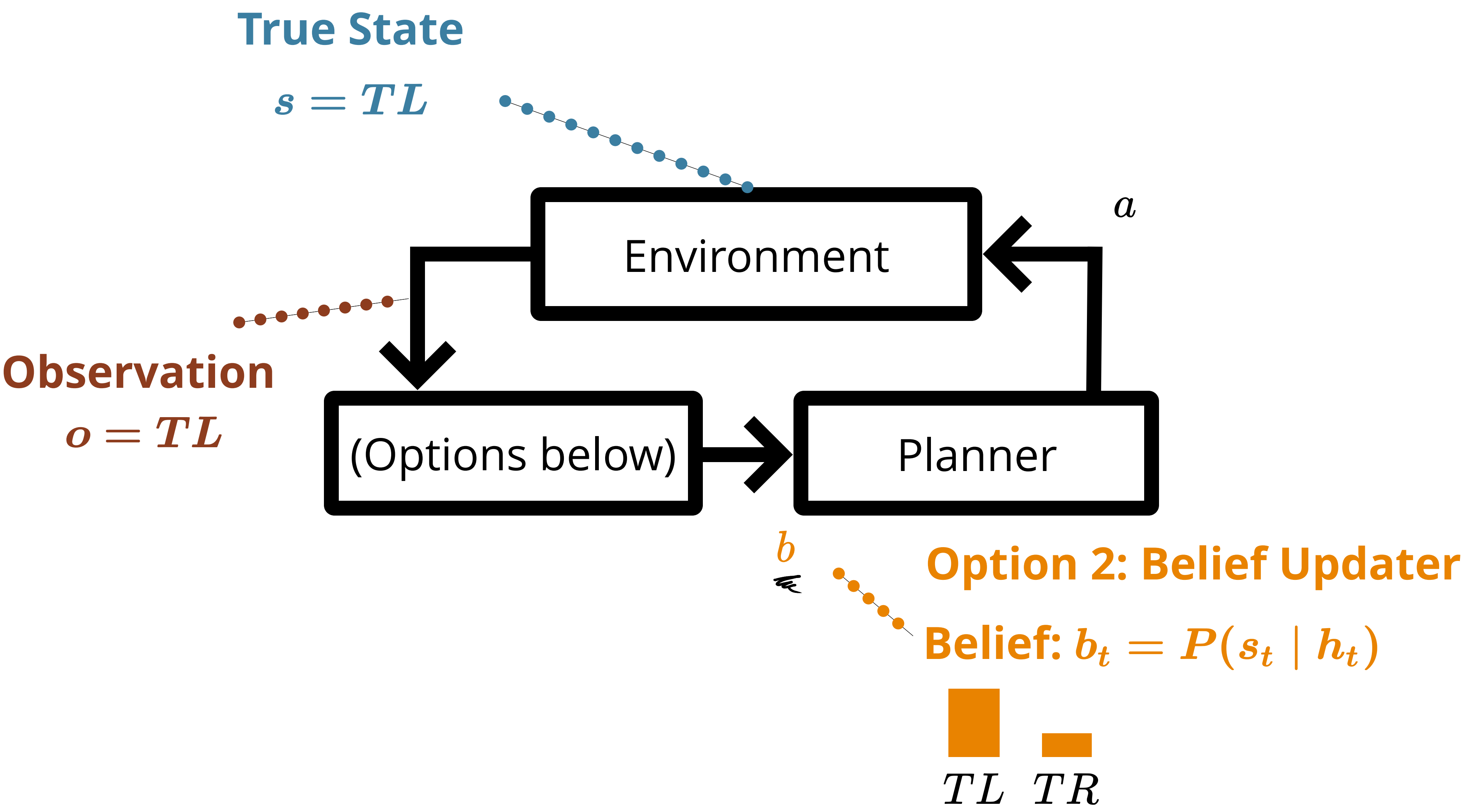
(solve a slightly different problem)

Last Time

POMDP Sense-Plan-Act Loop

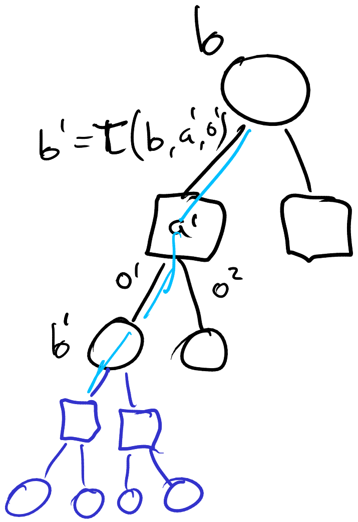


POMDP Sense-Plan-Act Loop



Belief-Space Tree Search: AEMS

G



while time remains

→ $b^* = \arg\max_{b \in \text{fringe}(G)} E(b)$

→ expand(b^*)

backup(b^*)

$$E(b) = \gamma^d P(b^d)\, \hat{\varepsilon}(b^d)$$ ← problem-specific

$$\hat{\varepsilon}(b) = U(b) - L(b)$$

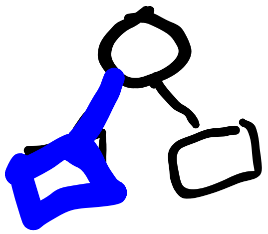
$$P(b^d) = \prod_{i=0}^{d-1} \underbrace{P(o^i \mid b^i, a^i)}_{T,Z} P(a^i \mid b^i)$$

$$P(a \mid b) = \frac{U(a,b) - L(b)}{U(b) - L(b)}$$ AEMS 1

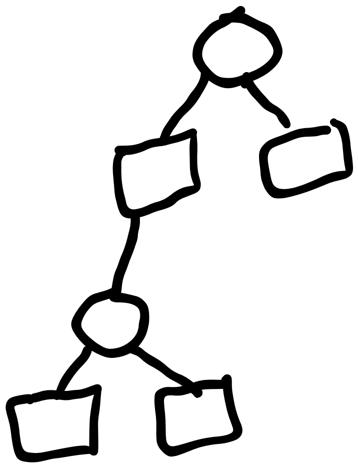
$$P(a \mid b) = \begin{cases} 1 & \text{if } a = \arg\max_{a'} U(a', b) \\ 0 & \text{o.w.} \end{cases}$$ AEMS 2

Monte Carlo Tree Search (MCTS/UCT)

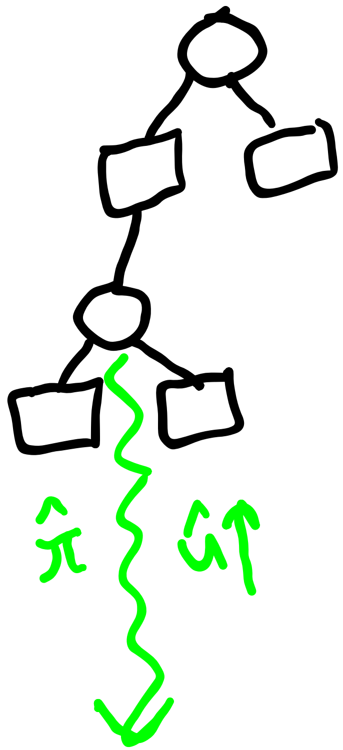
Search



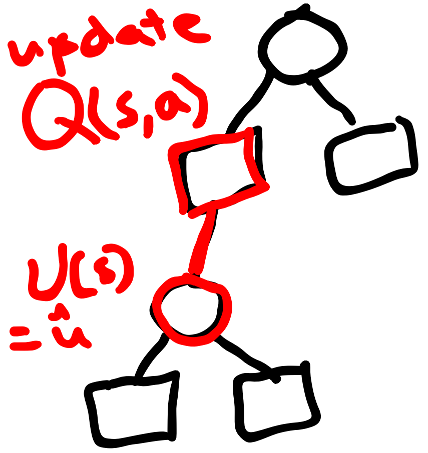
Expansion



Rollout



Backup

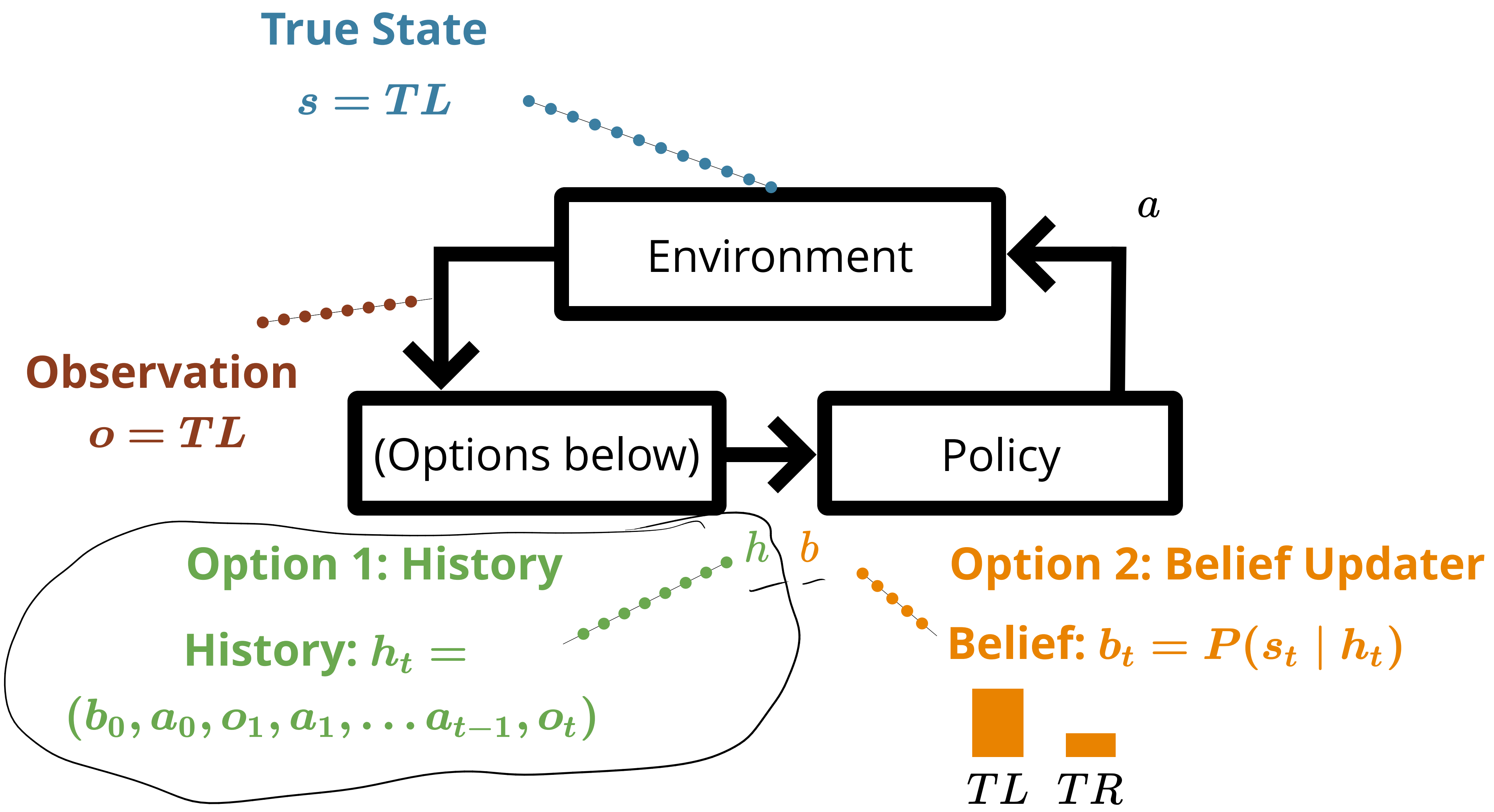


$$\underbrace{Q(s,a) + c\sqrt{\frac{\log N(s)}{N(s,a)}}}$$

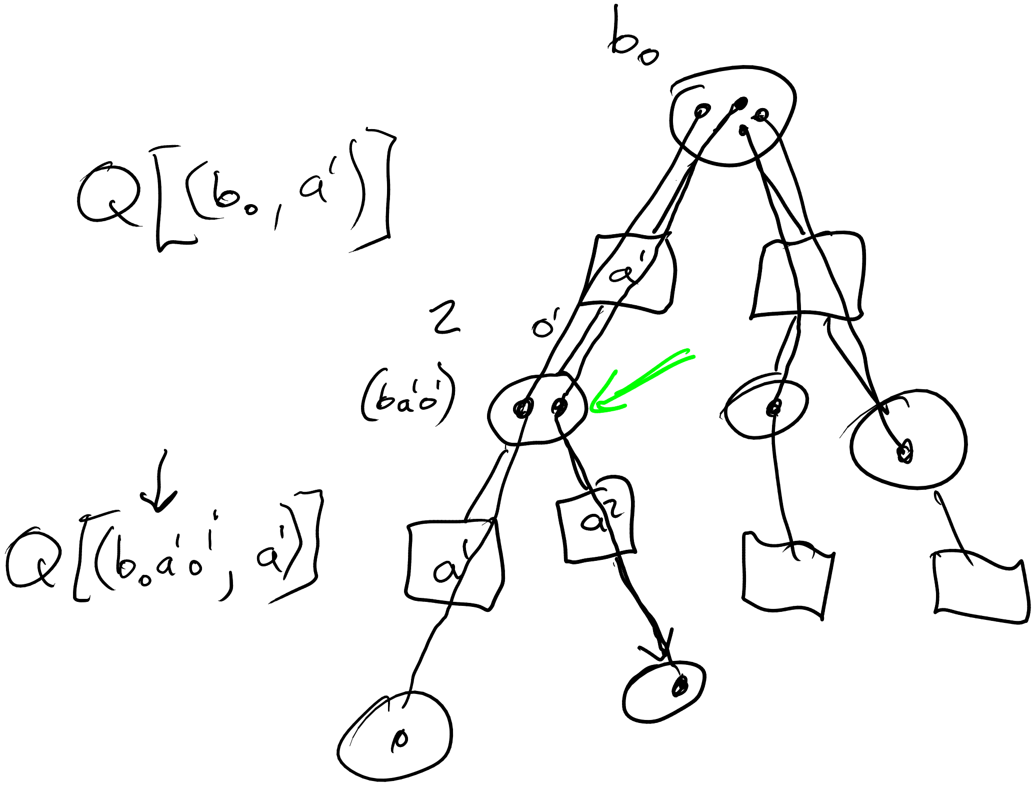
low $N(s,a)/N(s)$ = high bonus

start with $c = 2(\bar{V} - \underline{V})$

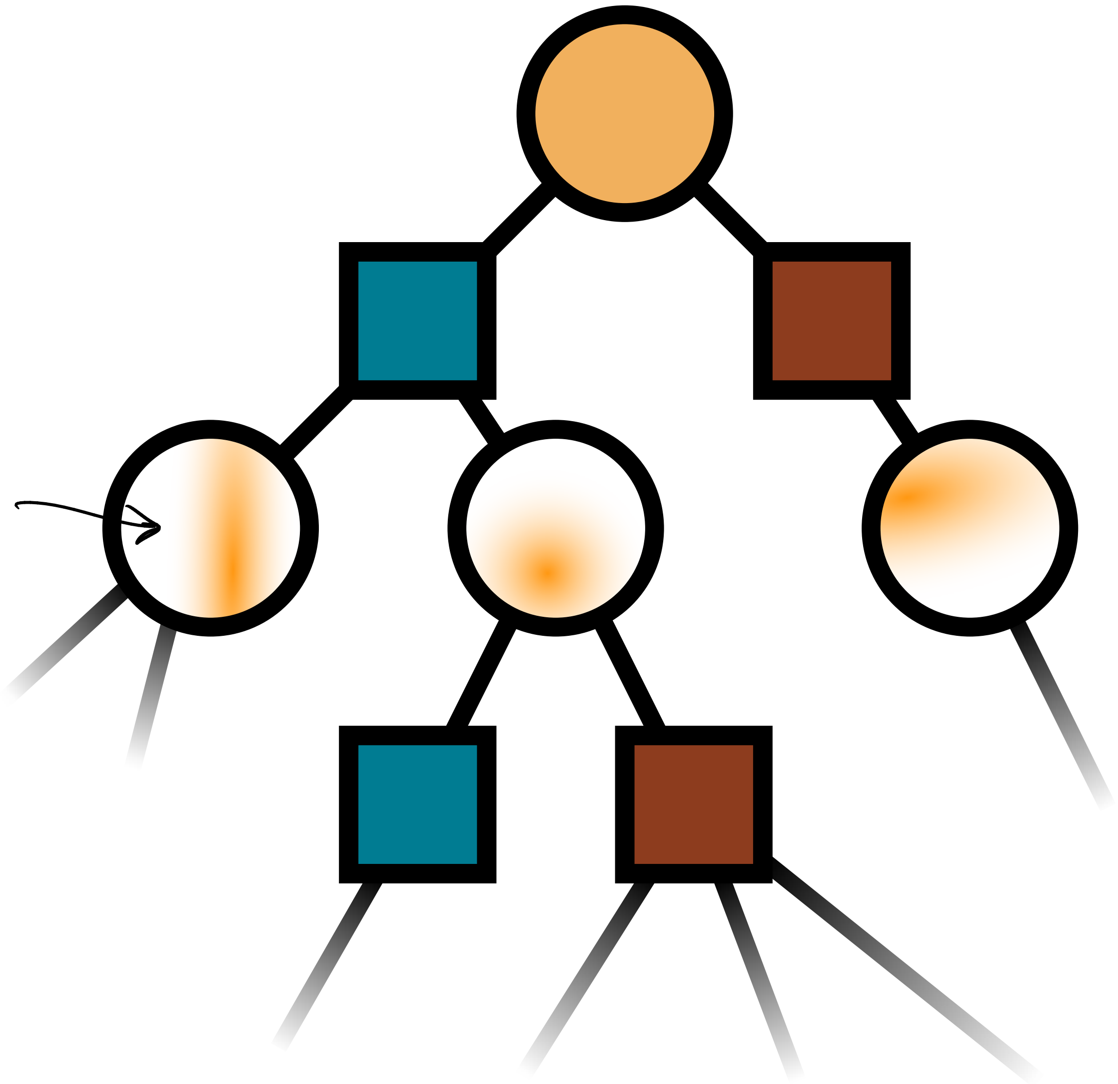
How should we adapt MCTS for POMDPs?

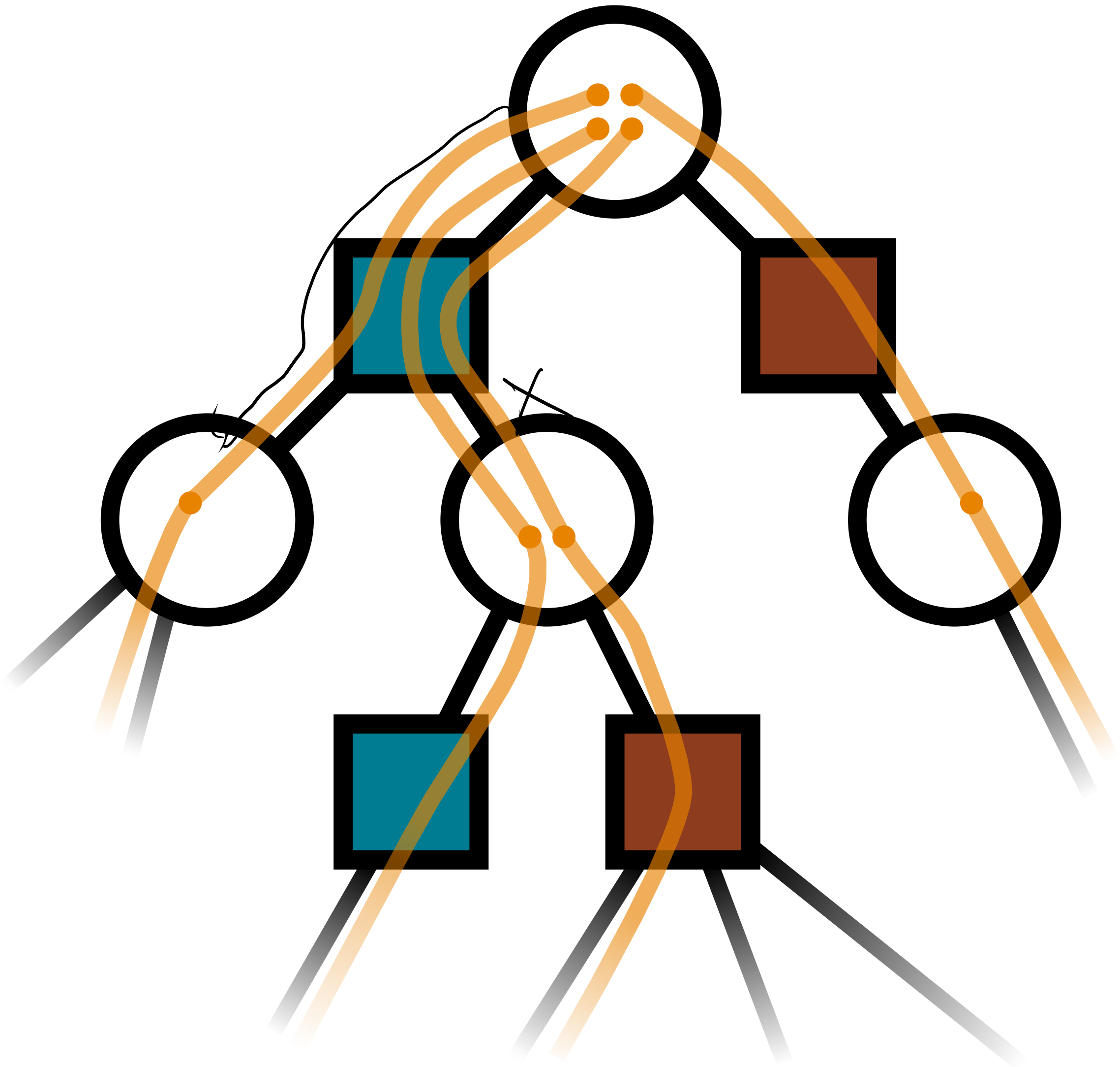


MCTS on Histories

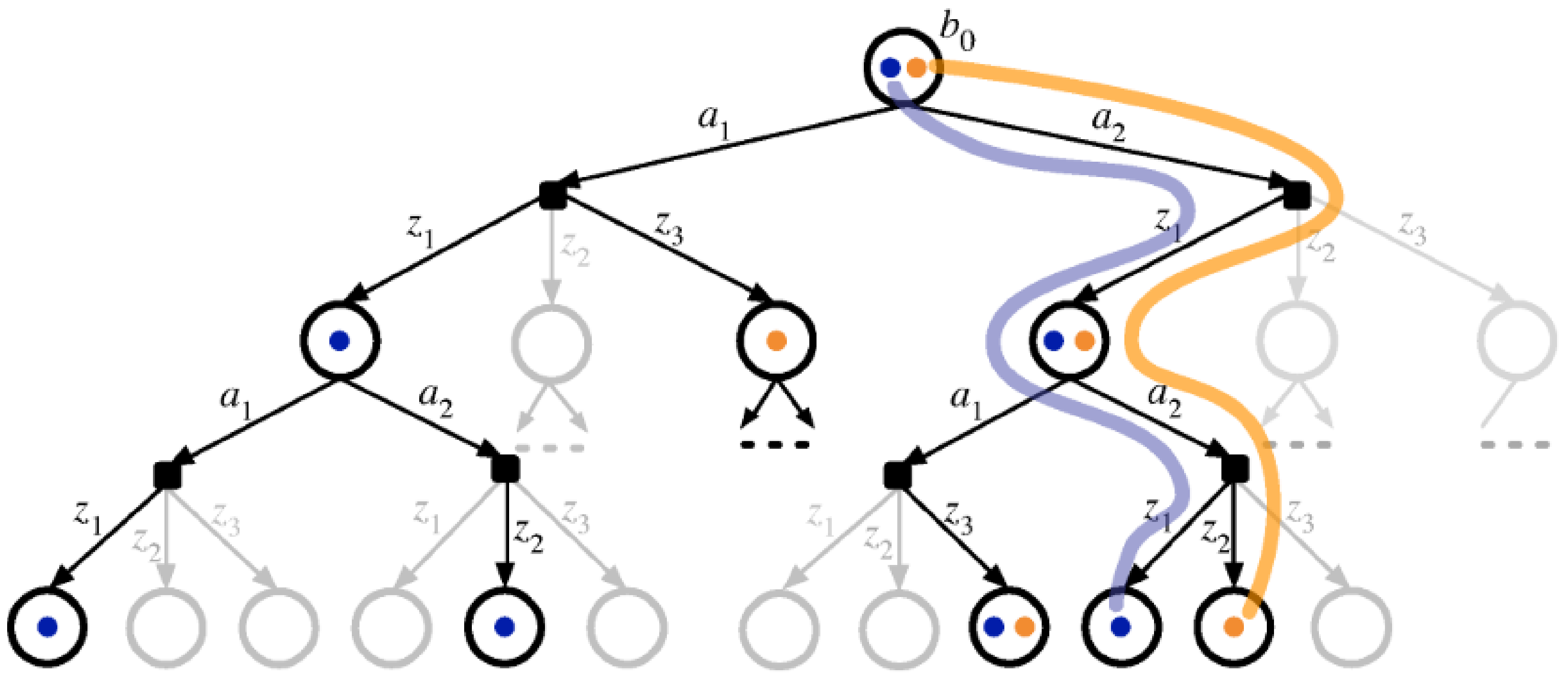


$(node, a, o) \longrightarrow$ next node



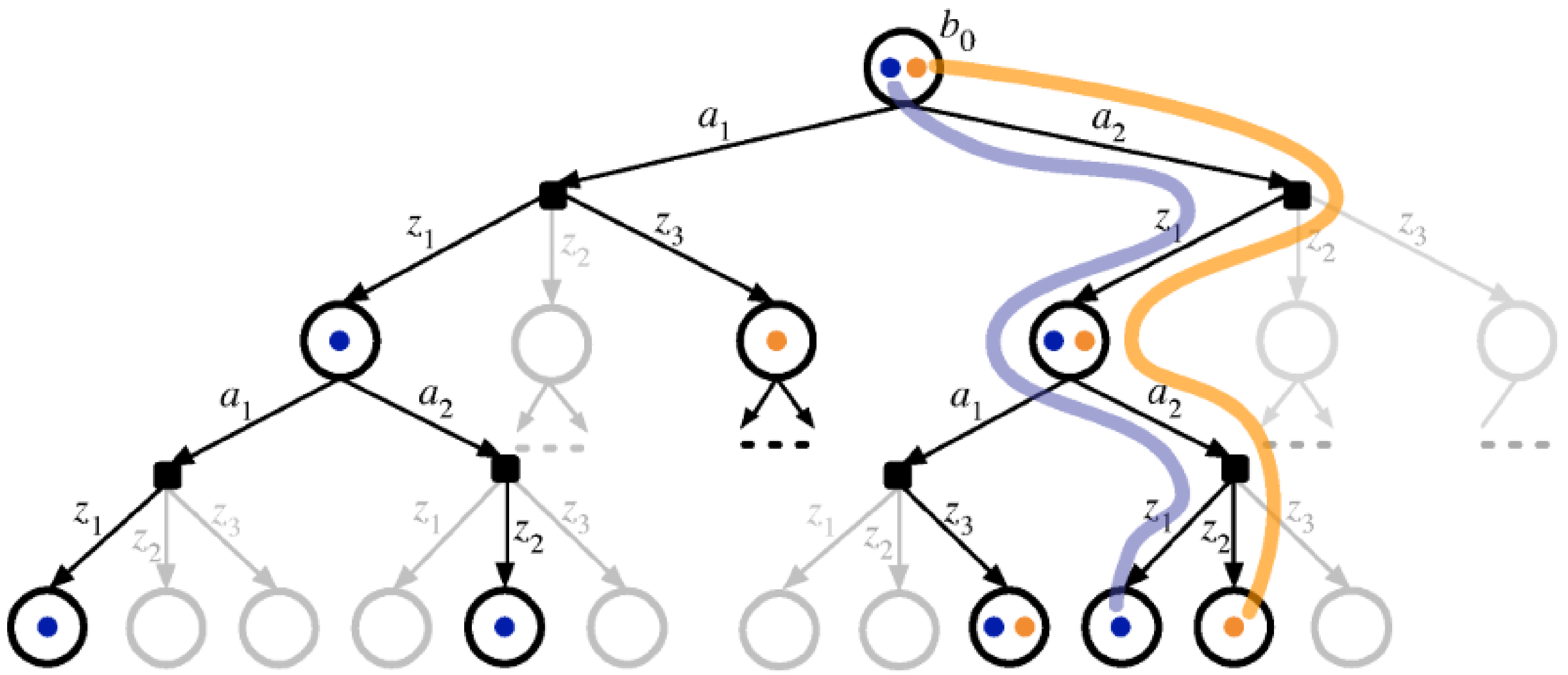


DESPOT



Somani, A., Ye, N., Hsu, D., & Lee, W. "DESPOT : Online POMDP Planning with Regularization." *Journal of Artificial Intelligence Research*, 2017

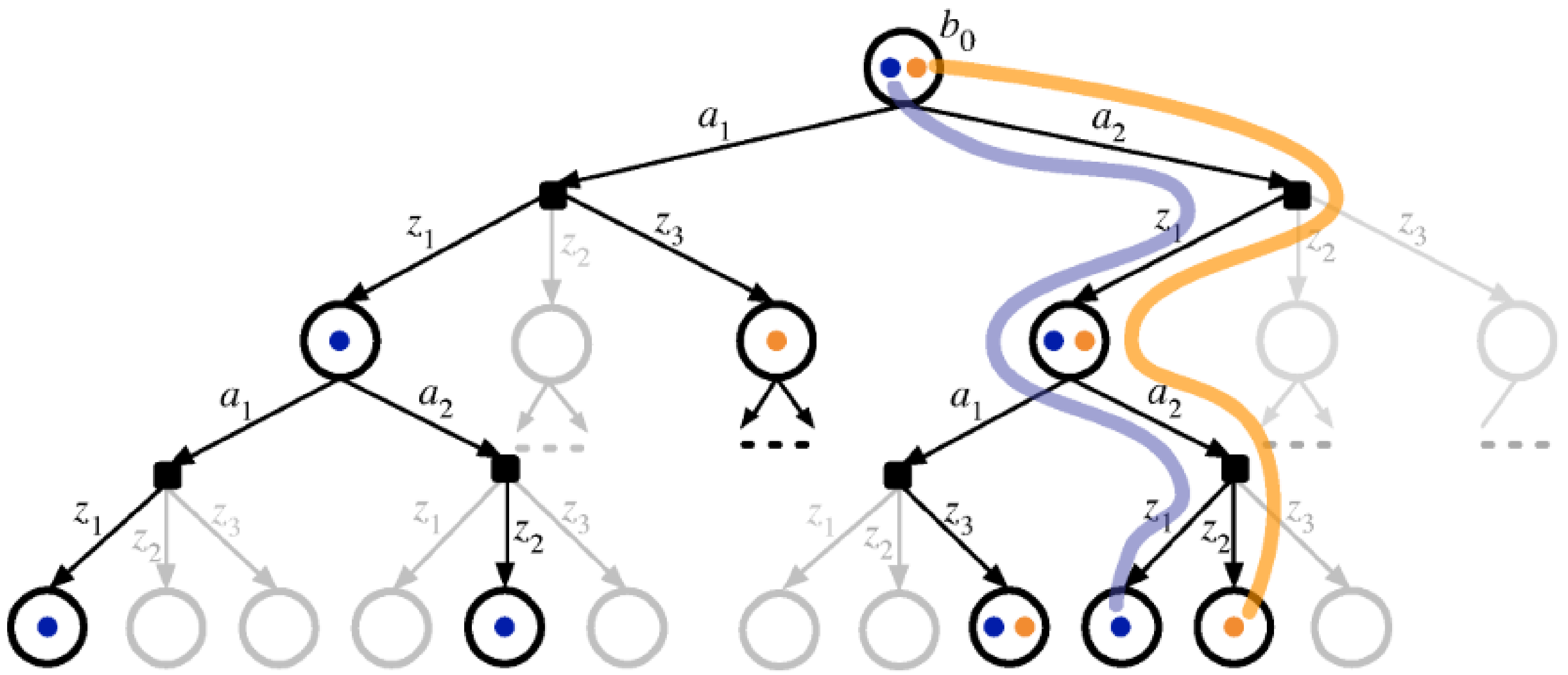
DESPOT



- Determinized Scenarios

Somani, A., Ye, N., Hsu, D., & Lee, W. "DESPOT : Online POMDP Planning with Regularization." *Journal of Artificial Intelligence Research*, 2017

DESPOT

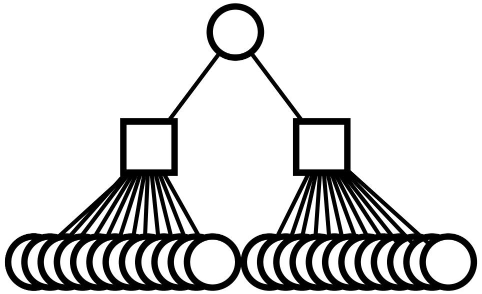


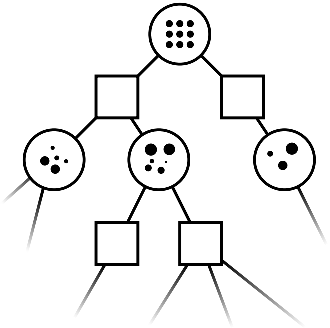
- Determinized Scenarios
- Guided by Lower and Upper Bounds

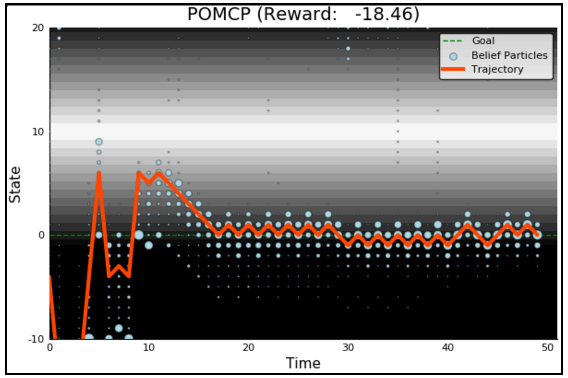
Somani, A., Ye, N., Hsu, D., & Lee, W. "DESPOT : Online POMDP Planning with Regularization." *Journal of Artificial Intelligence Research*, 2017

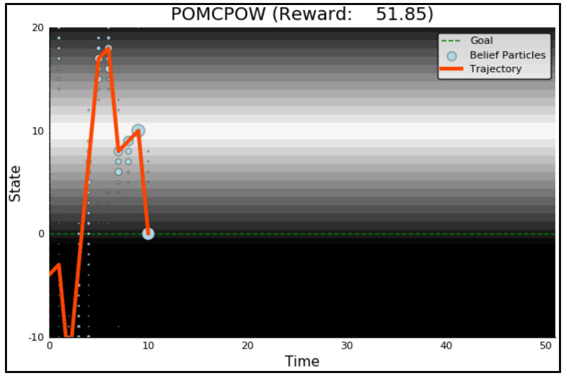
POMCP

POMCPOW





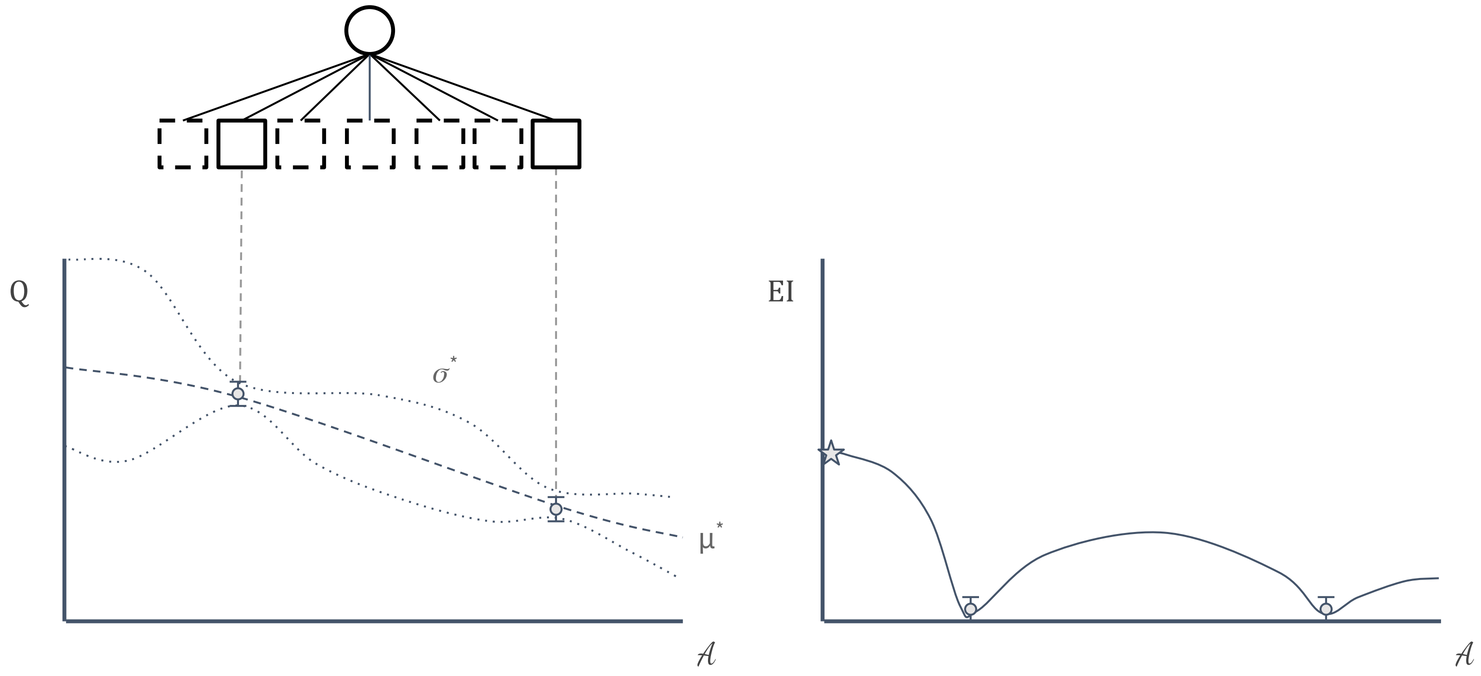




DESPOT-α

BOMCP

Bayesian Optimized Action Branching



[Mern, Sunberg, et al. AAAI 2021]

BOMCP

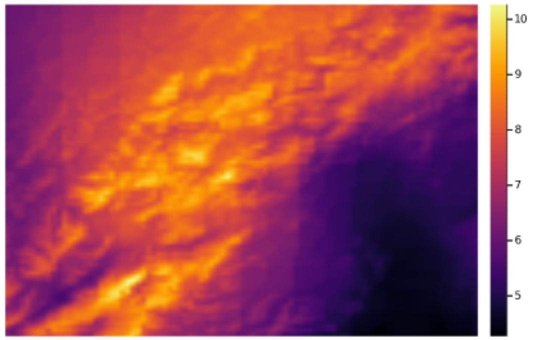
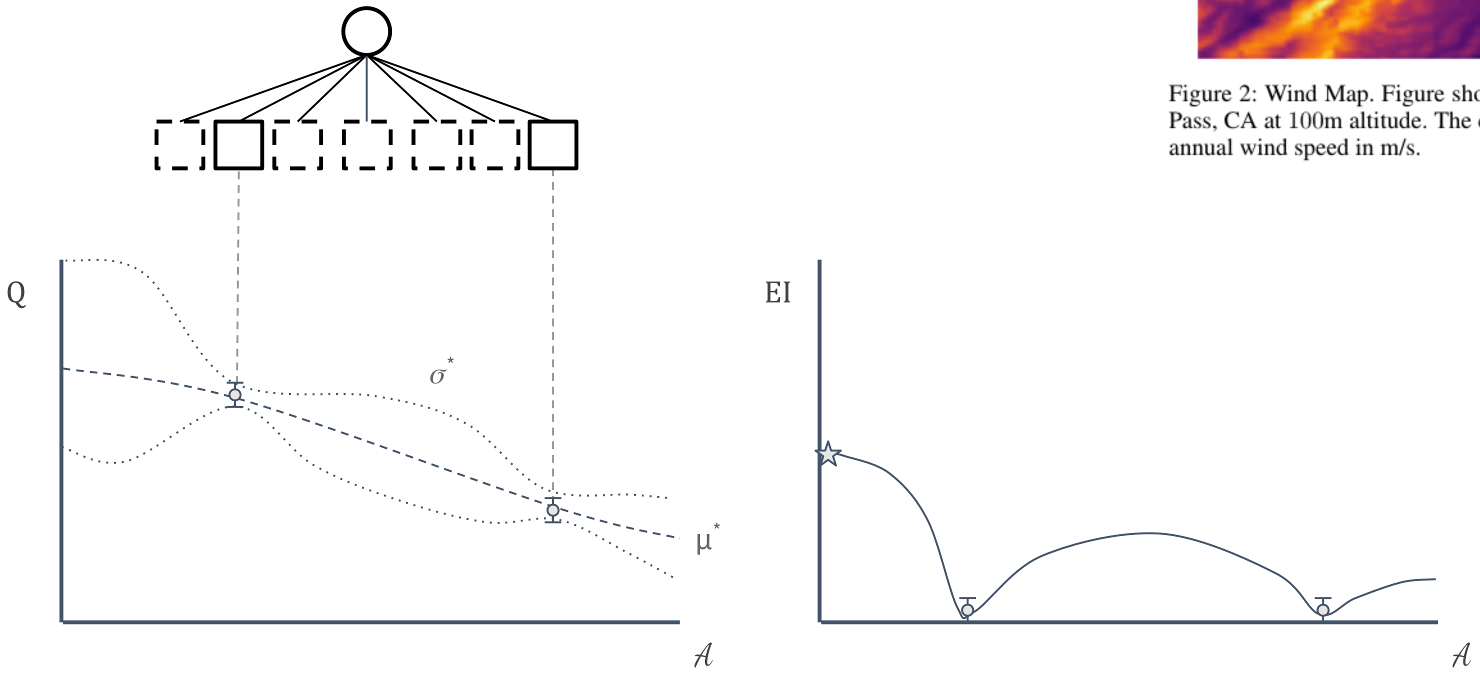


Figure 2: Wind Map. Figure shows wind map for Altamont Pass, CA at 100m altitude. The colors represent the average annual wind speed in m/s.

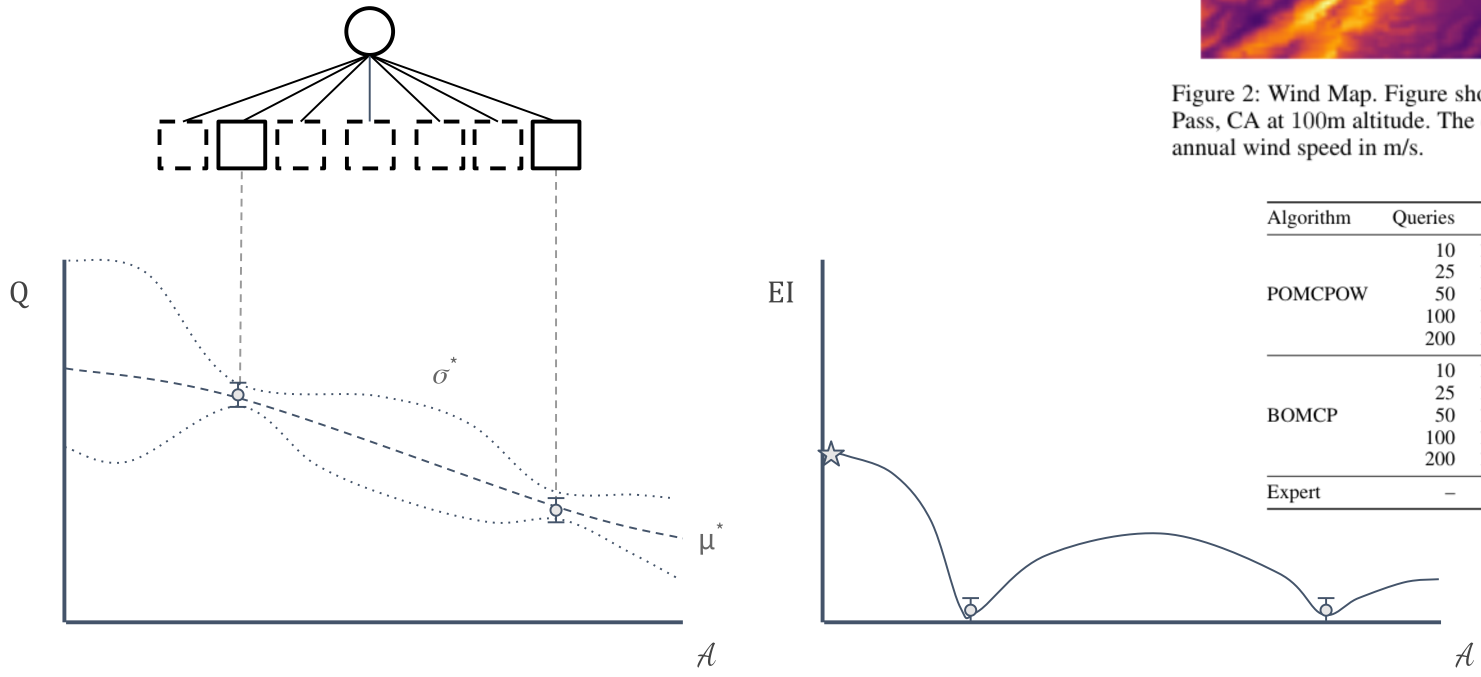
Bayesian Optimized Action Branching



[Mern, Sunberg, et al. AAAI 2021]

BOMCP

Bayesian Optimized Action Branching



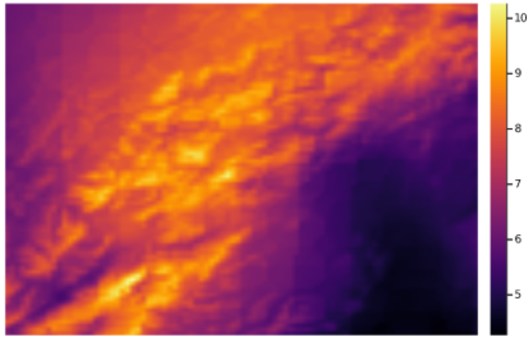
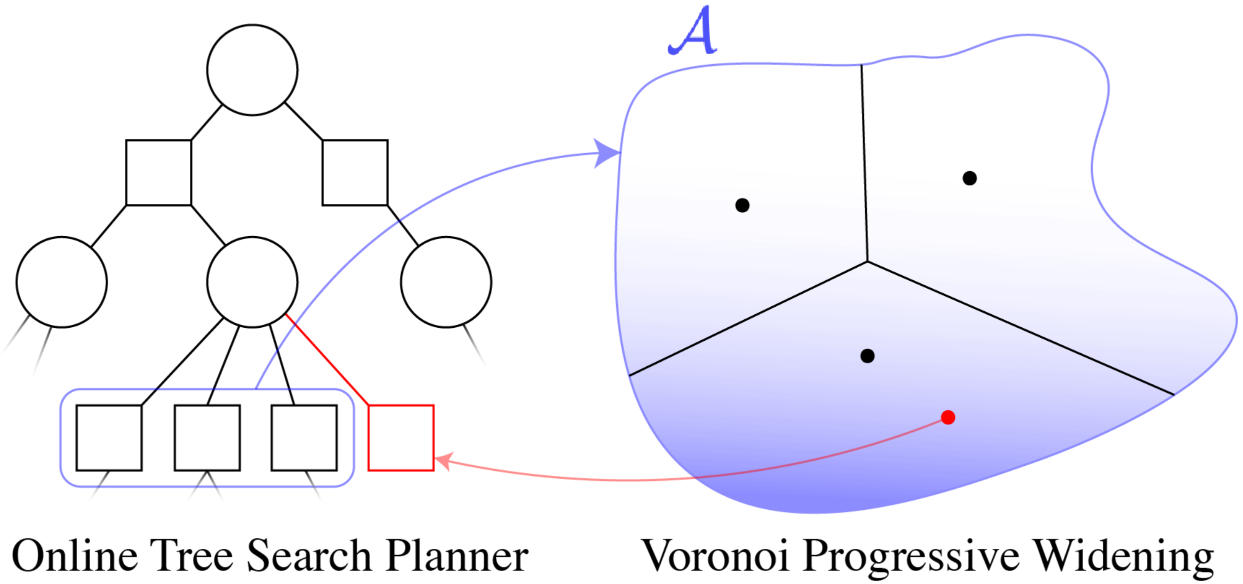


Figure 2: Wind Map. Figure shows wind map for Altamont Pass, CA at 100m altitude. The colors represent the average annual wind speed in m/s.

Algorithm	Queries	Score	Time (seconds)
POMCPOW	10	15708 ± 229	2.25 ± 0.07
	25	16234 ± 217	4.80 ± 0.07
	50	16374 ± 212	6.27 ± 0.08
	100	16018 ± 262	11.98 ± 0.07
	200	15787 ± 233	20.67 ± 0.09
BOMCP	10	18095 ± 183	2.55 ± 0.08
	25	18154 ± 158	5.21 ± 0.07
	50	18015 ± 163	6.71 ± 0.06
	100	18225 ± 119	13.39 ± 0.07
	200	18113 ± 157	25.14 ± 0.08
Expert	–	8130 ± 51	–

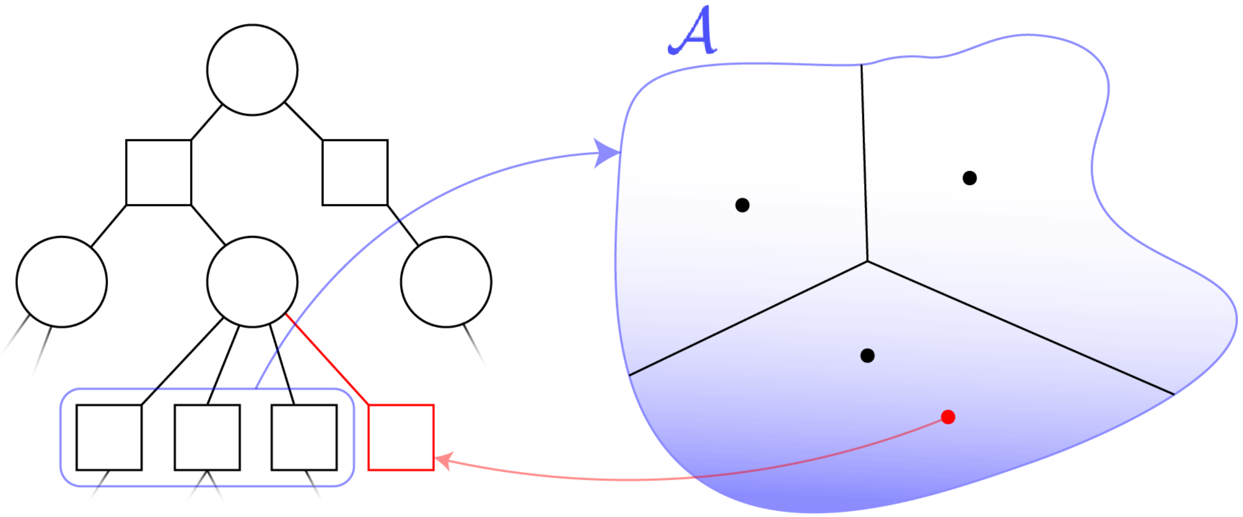
[Mern, Sunberg, et al. AAAI 2021]

Voronoi Progressive Widening



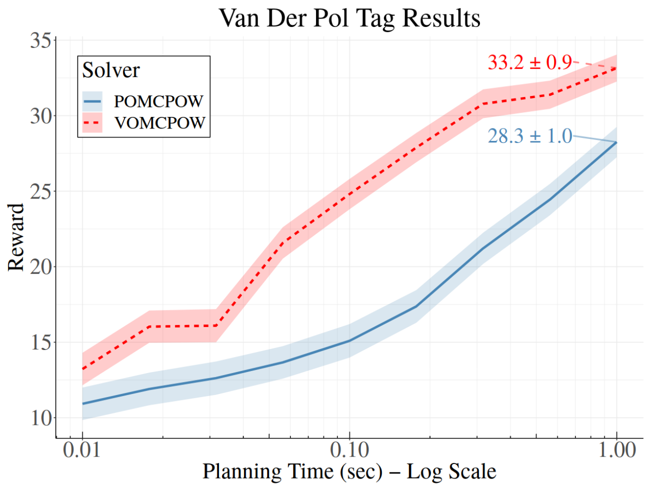
[Lim, Tomlin, & Sunberg CDC 2021]

Voronoi Progressive Widening



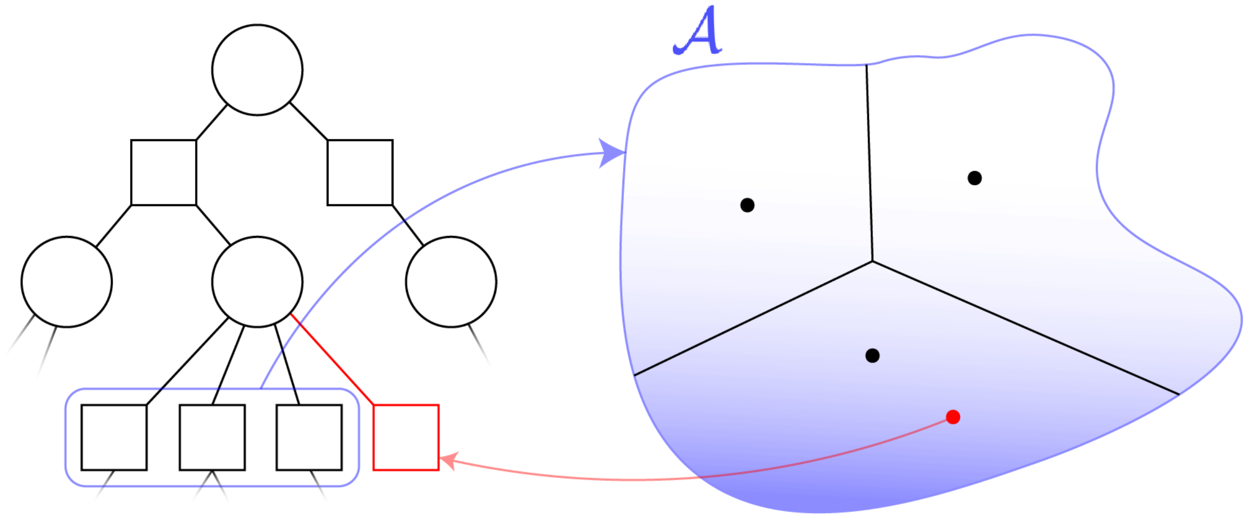
Online Tree Search Planner

Voronoi Progressive Widening

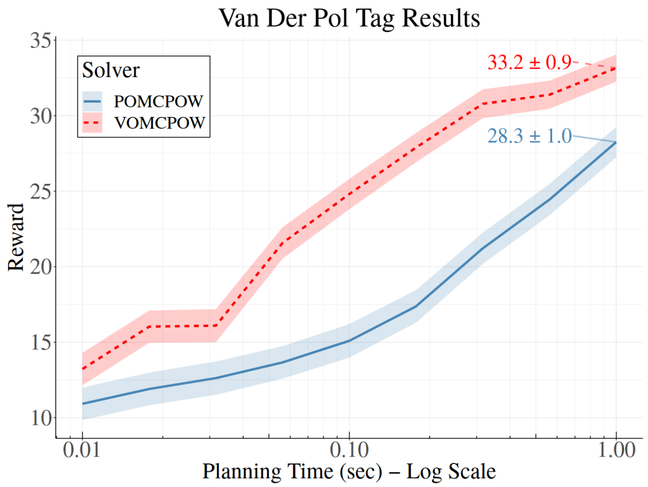


[Lim, Tomlin, & Sunberg CDC 2021]

Voronoi Progressive Widening



Online Tree Search Planner   Voronoi Progressive Widening



Theorem 2 (VOWSS Inequality)**.** *Given the action sampling width of C_a and state sampling width of C_s at every height of the tree that follow the intermediate concentration bounds in the form of POWSS (Lim, Tomlin, and Sunberg 2020) and regret bounds in the form of VOO (Kim et al. 2020), the following bounds for the VOWSS estimator $\hat{V}^{C_a}_{\text{VOWSS},d}(b)$ hold for all $d \in [0, D-1]$ in expectation:*

$$\left| V^{\star}_{d}(b) - \hat{V}^{C_a}_{\text{VOWSS},d}(b) \right| \leq \eta + \alpha \qquad ($$

[Lim, Tomlin, & Sunberg CDC 2021]