Information-Lookahead Planning for AUV Mapping

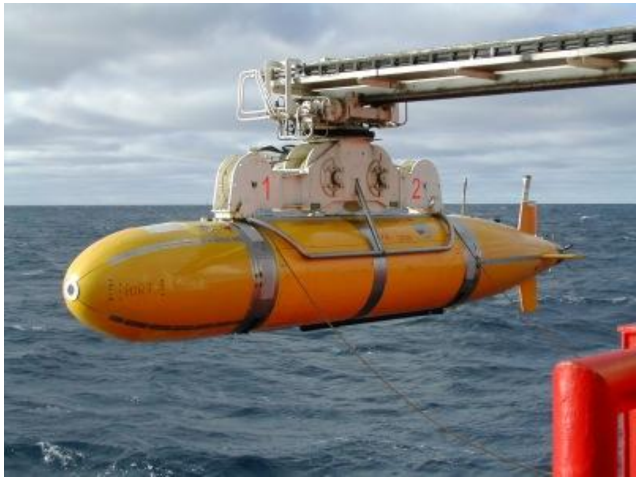
Zeyn Saigol[1], Richard Dearden[1],
Jeremy Wyatt[1] and Bramley Murton[2]

[1]University of Birmingham, UK
[2]NOC Southampton, UK

IJCAI 2009

Outline

1. Oceanography domain
2. Problem formalisation
3. Information-lookahead algorithm
4. Results
5. Certainty-equivalent heuristic

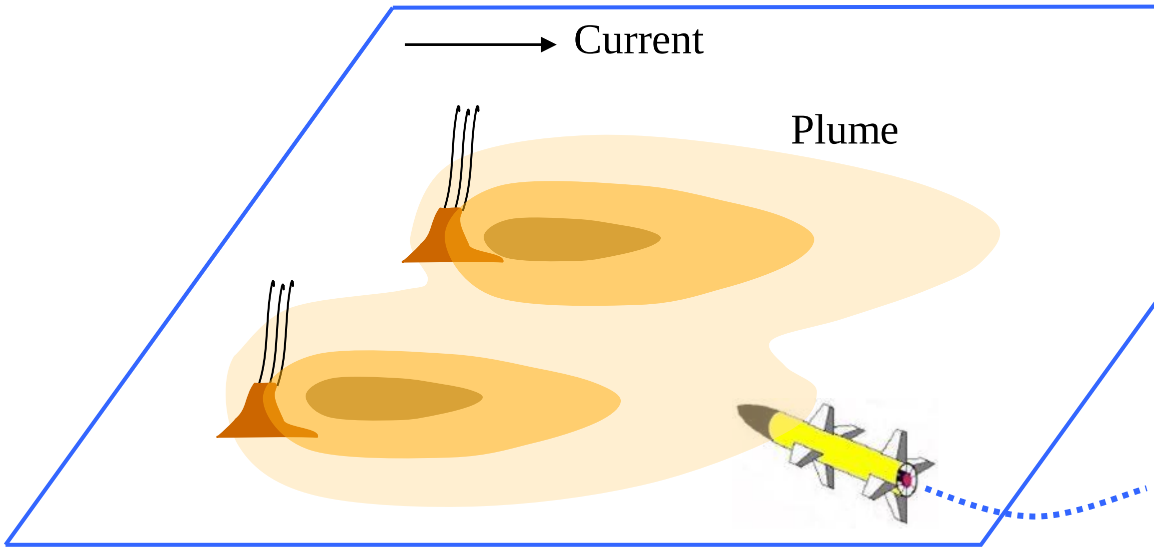


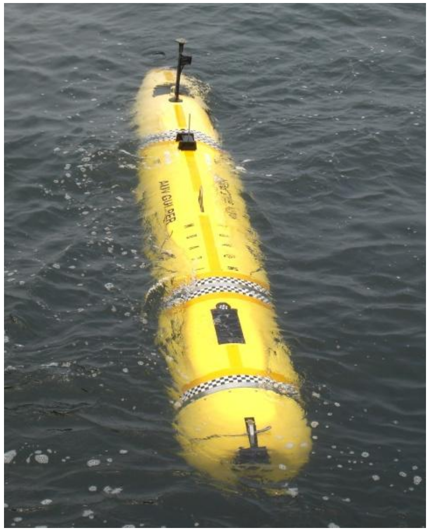
Introduction

- Planning to optimise robotic scientific exploration.
- Large, real-world problem: significant contribution is to formalise as a POMDP.
- Oceanography domain – autonomous underwater vehicle (AUV).
- Task is to hunt for *hydrothermal vents*, found on the sea floor.

Planning Problem

- Hydrothermal vent prospecting
- Current methods require exhaustive "mow-the-lawn" survey



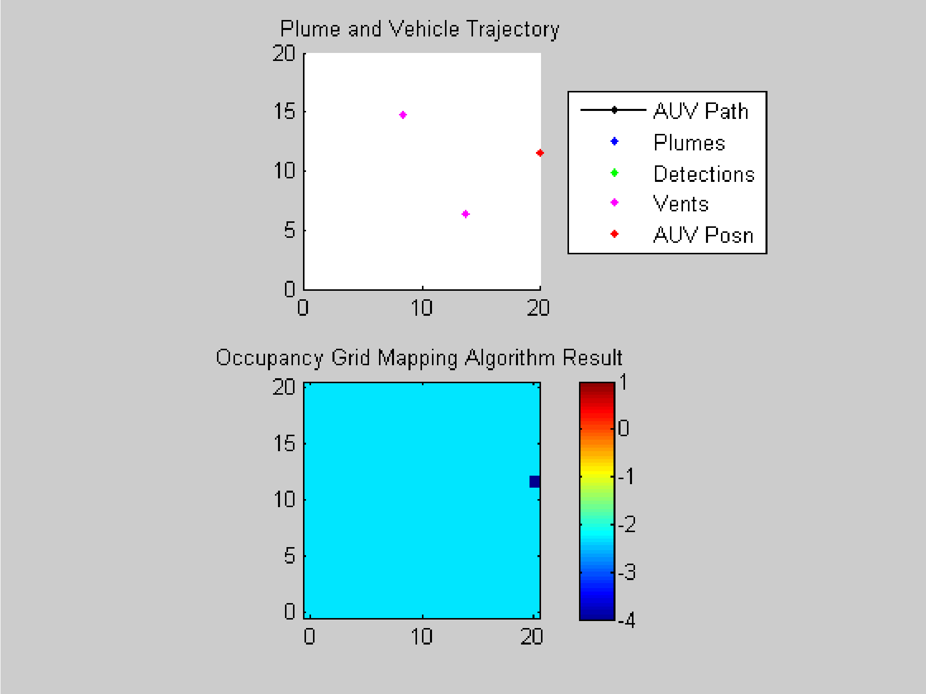


Planning challenges:

- Limited mission duration
- Unknown number of vents
- Unknown vent locations
- Imprecise observation model
- Continuous state space

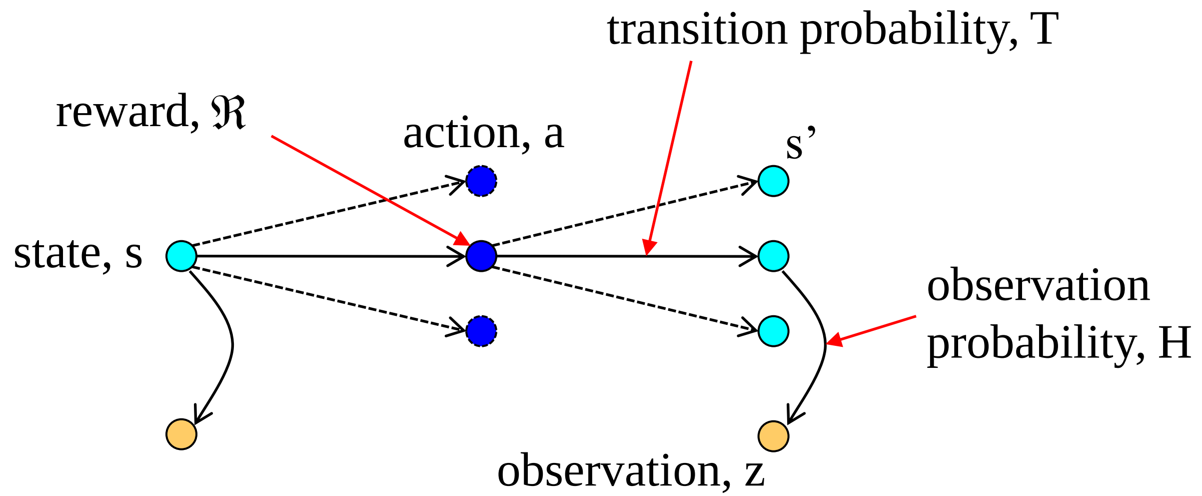
Mapping

- Split the problem into two phases:
  1. Mapping: given sequence of observations, create a probabilistic map of most likely vent locations.
  2. Planning: given the uncertain map of vents, decide on a strategy for finding the maximum number of vents.
- Mapping is handled by the occupancy grid method of Michael Jakuba (Jakuba, 2007)
- Creates a map of probability of each cell containing a vent, $P(m_c)$



POMDP Recap

- POMDPs are an extension of Markov decision processes (MDPs) to partial observability.
- Defined by $\langle S, A, T, \Re, Z, H \rangle$



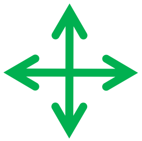
- Agent can act optimally by tracking its *belief state*, b, which is a distribution $P(s)$ over possible states.
- Any POMDP can be converted to an *information-state MDP (ISMDP)* where belief states become states in the ISMDP.

Problem Model

Belief states

- Cell AUV is in, ocean current, list of already-found vents (known)
- Occupancy grid

Actions

- Move to adjacent grid cells 

Transition Function

- Observations $z = \{l, p\}$

  l: find vent (avoid multiple rewards), p: detect plume
- Observation model $P(z \mid b, a)$ derived using same plume model used by Jakuba's OG update (see paper)
- Observation model together with OG update gives $P(b' \mid b, a, z)$

Reward

$$\Re(b,a) = \begin{cases} P(m_c)R_{vent} & \text{if c is an unvisited cell} \\ 0 & \text{otherwise} \end{cases}$$

Previous Work

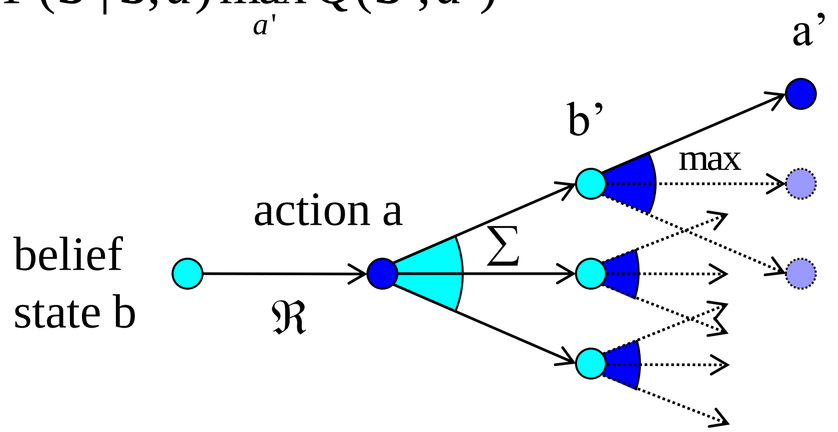
- Exact POMDP solvers can handle problems with only dozens of states (Kaelbling et. al, 1998).
- Point-based approximations such PBVI (Pineau et. al., 2003) and HSVI (Smith and Simmons, 2004) increased this to $\sim 10^5$ states.
- For this domain, with a grid of 100 by 100 cells, we have 10^{3000} states.
- Online methods (Ross et. al., 2008) are much closer to being able to handle this domain
  - Work by searching forward in belief space from a start belief state.
  - Need to solve a relaxed version of the POMDP to assign values to leaf nodes => still need a value for every state.
- Also research on chemical plume tracing (Vergassola et. al., 2007; Russell et. al., 2003), but this is only applicable to single-source problems.

IL Algorithm (1)

Information-Lookahead calculation

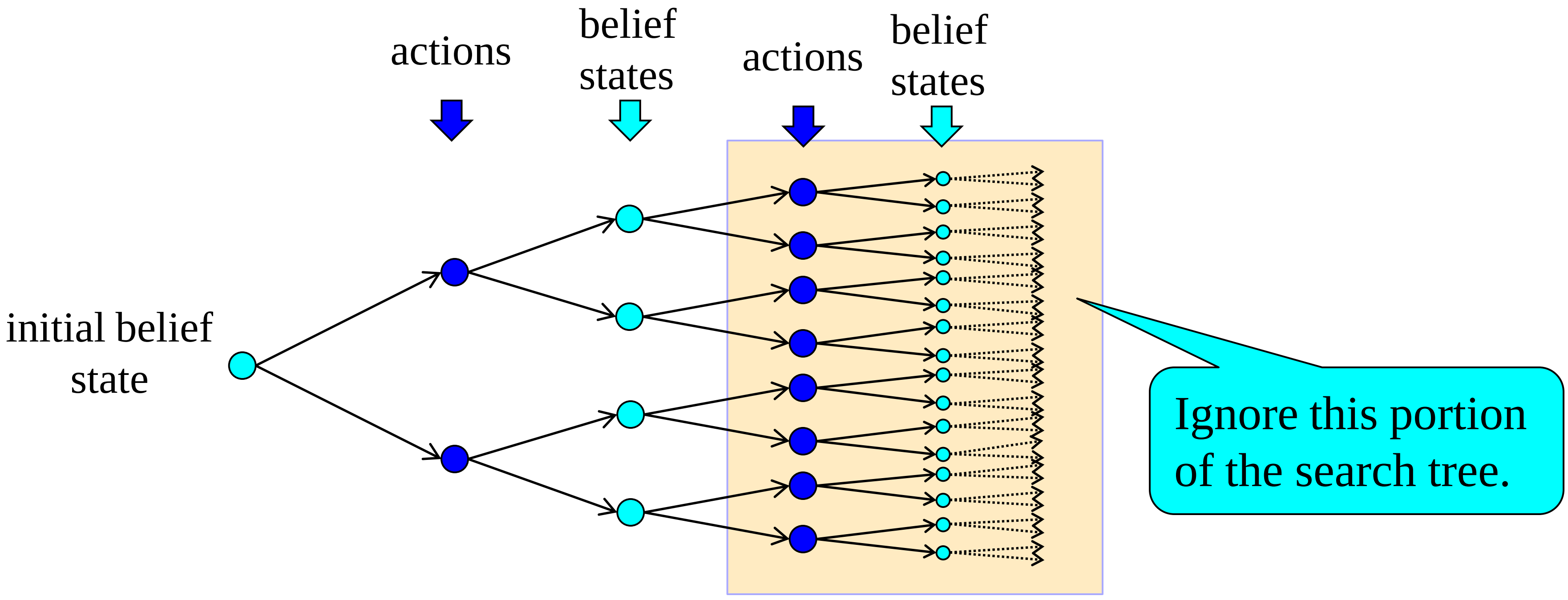
- Reward function means a value can be associated with each possible future action/outcome path.
- Use standard dynamic programming to propagate values back from leaf nodes to the current belief state (c.f. reinforcement learning).
- Note this requires running the OG update to calculate each new b’.

$$Q(\mathbf{b},a) = \Re(\mathbf{b},a) + \gamma \sum_{\mathbf{b}'} P(\mathbf{b}' \mid \mathbf{b},a) \max_{a'} Q(\mathbf{b}',a')$$



IL Algorithm (2)

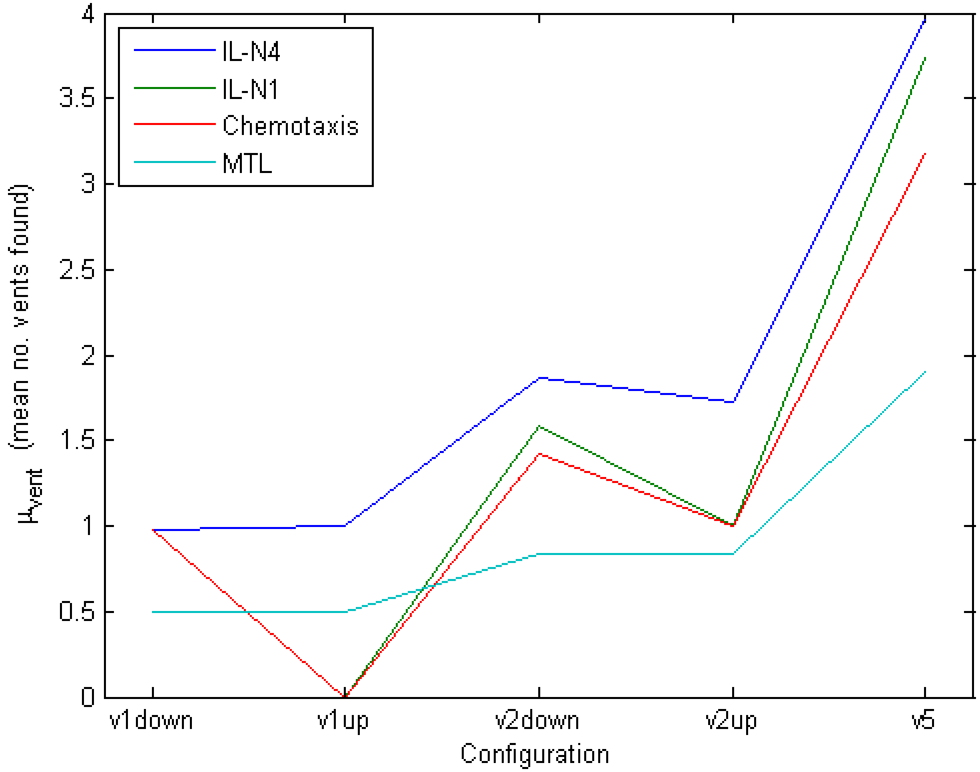
- Exponential in the number of actions taken – so intractable to plan for whole mission duration.
- Approximate with a small number of action steps (i.e. pretend it's a finite horizon problem with a very small horizon).



Results

Algorithms were:

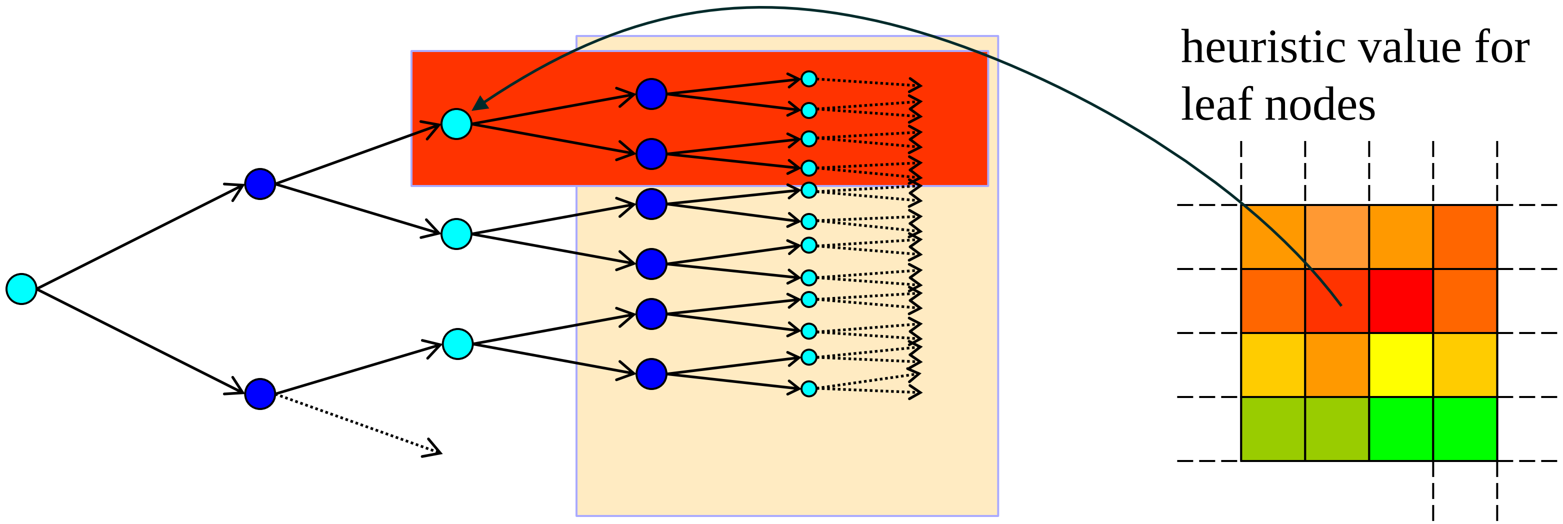
- IL with four steps lookahead
- IL with one step of lookahead
- Chemotaxis – moth-like behaviour
- Mowing-the-lawn



Experiments were run on 20x20 grids with five different arrangements of vents, with 50 trials of each.

Certainty-Equivalent Extension

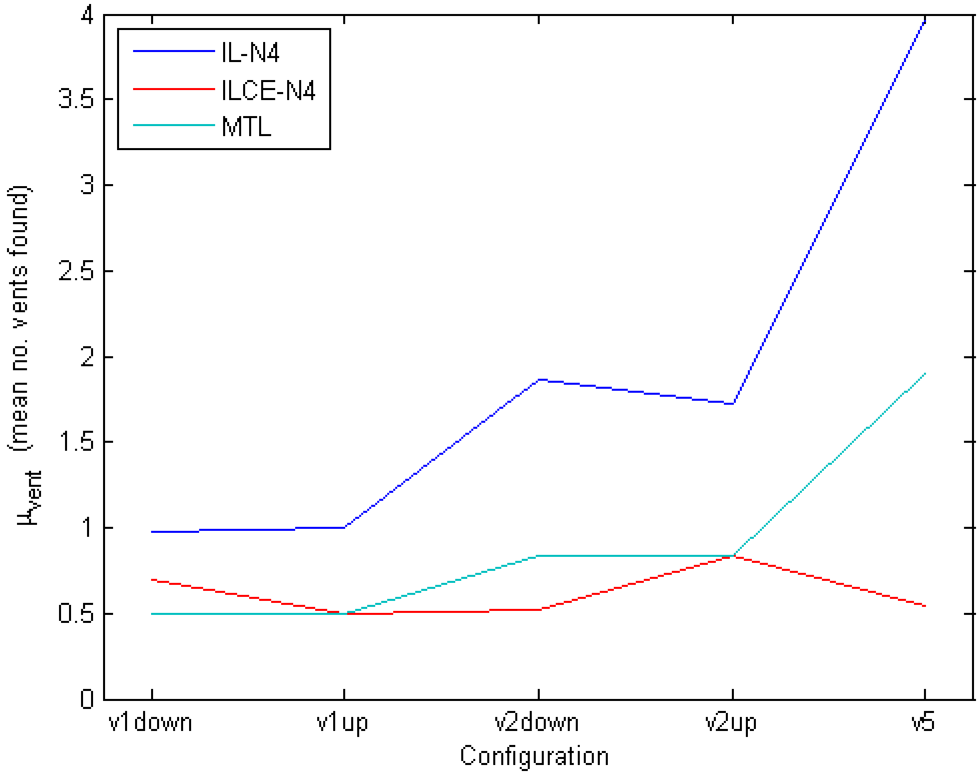
- Instead of just ignoring the future after N steps, use a heuristic estimate of its expected value.



CE (certainty-equivalent)

- Deterministic MDP in belief-state space (very fast to solve).
- Observations are ignored, i.e. cell occupancy probabilities are fixed.
- List of visited vents is dropped from the state space.

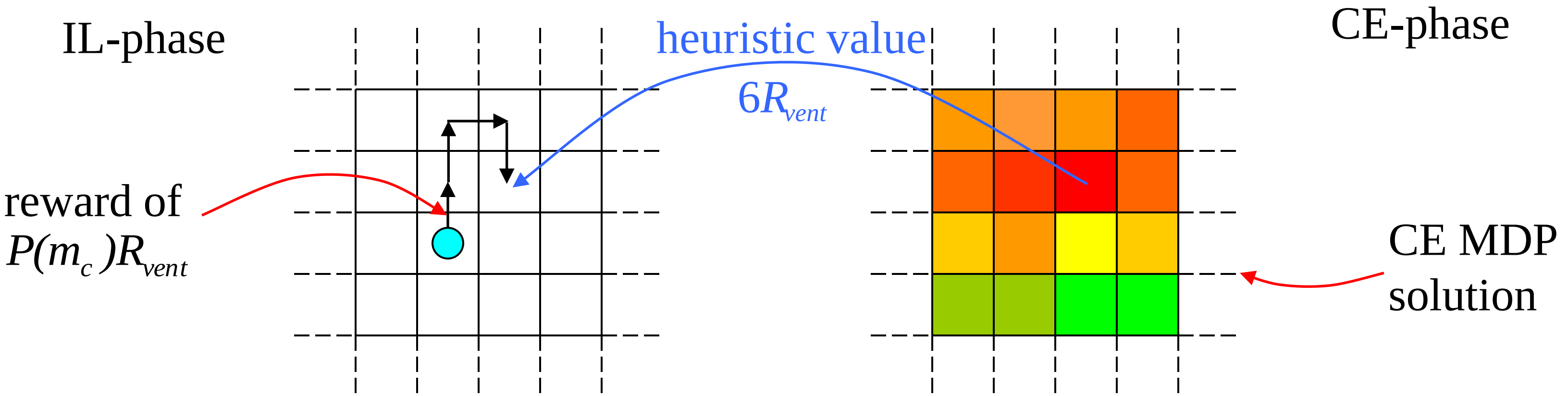
CE Results



Information-lookahead with the CE-heuristic performed even worse than mowing-the-lawn!

CE Failure

- CE values are calculated using dynamic programming, so cell probabilities are not zeroed once they have been visited.
- This means cell values can grow much larger than $P(m_c)R_{vent}$, as high-value cells can be re-visited repeatedly .



- Therefore the agent gets more reward by delaying visiting high-value cells to the CE-phase; as the visit is again delayed on the next step, high-value cells are actively avoided by the IL-CE algorithm.

Conclusions

- Interesting real-world application for AI planning.
- Large state space is a real challenge for state-of-the-art POMDP solvers.
- Formally defined the problem as a POMDP.
- Information-lookahead solution using forward search from a start belief state outperforms existing methods (in simulation).
- Ideas for future work include:
  - Solving a set of hierarchical grids, where coarse grids inform the values for fine-grained grids near the agent's current location.
  - Using entropy minimisation techniques, which can avoid the exponential blowup in search space. However priors are << 0.5, so entropy increases when get useful data.
  - Deploy on a vehicle (longer term).

References

Jakuba, M. (2007). *Stochastic Mapping for Chemical Plume Source Localization with Application to Autonomous Hydrothermal Vent Discovery*. PhD thesis, MIT and WHOI Joint Program.

Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134.

Pineau, J., Gordon, G., and Thrun, S. (2003). Point-based value iteration: An anytime algorithm for POMDPs. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI).*

Ross, S., Pineau, J., Paquet, S., and Chaib-draa, B. (2008). Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*, 32:663 – 704.

Russell, R. A., Bab-Hadiashar, A., Shepherd, R. L., and Wallace, G. G. (2003). A comparison of reactive robot chemotaxis algorithms. *Robotics and Autonomous Systems*, 45(2):83–97.

Smith, T. and Simmons, R. (2004). Heuristic search value iteration for POMDPs. In *Proceedings of the 20th Annual Conference on Uncertainty in Articial Intelligence (UAI-04)*, pages 520–527, AUAI Press, Arlington, Virginia.

Vergassola, M., Villermaux, E., and Shraiman, B. I. (2007). 'Infotaxis' as a strategy for searching without gradients. *Nature*, 445(7126):406–409.