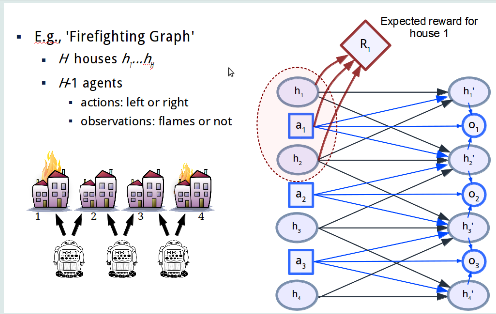
Approximate Solutions for Factored Dec-POMDPs with Many Agents

Frans A. Oliehoek, Shimon Whiteson, & Matthijs T.J. Spaan

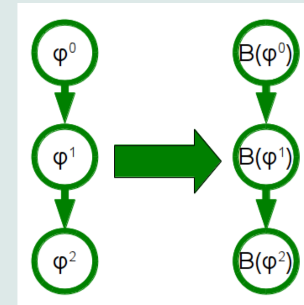
AAMAS, Wednesday May 8, 2013

Visual Roadmap

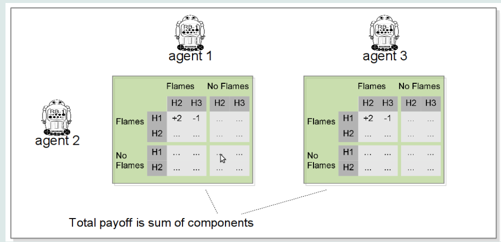
model



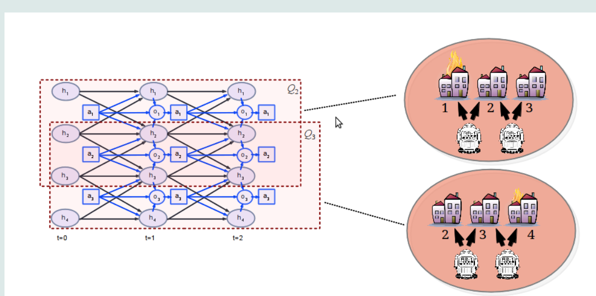
background: FSPC



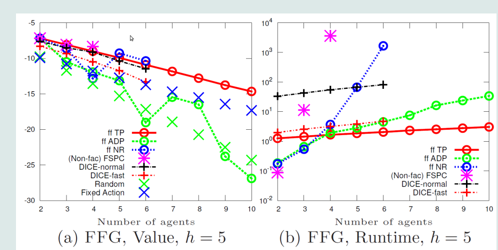
Factored FSPC



practical factored FFSPC

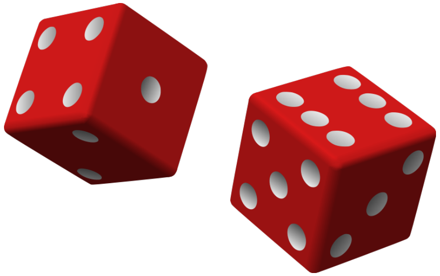


results

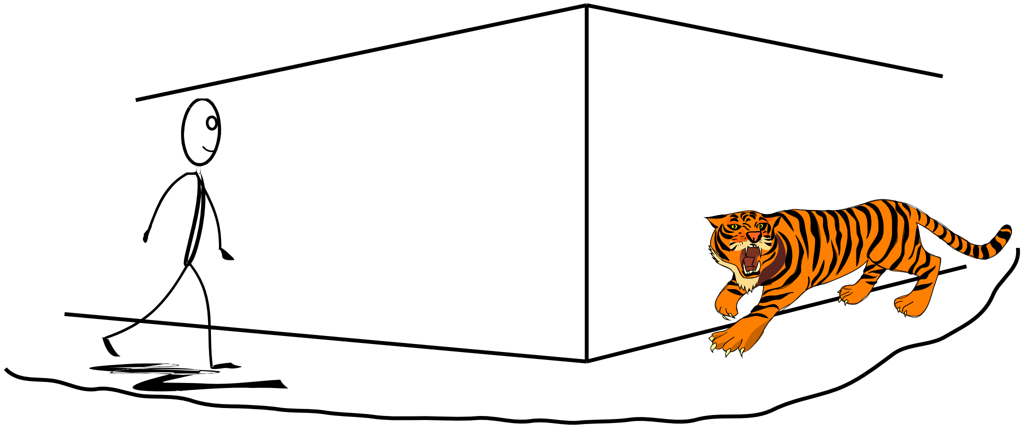


Multiagent decision making under Uncertainty

- Outcome Uncertainty

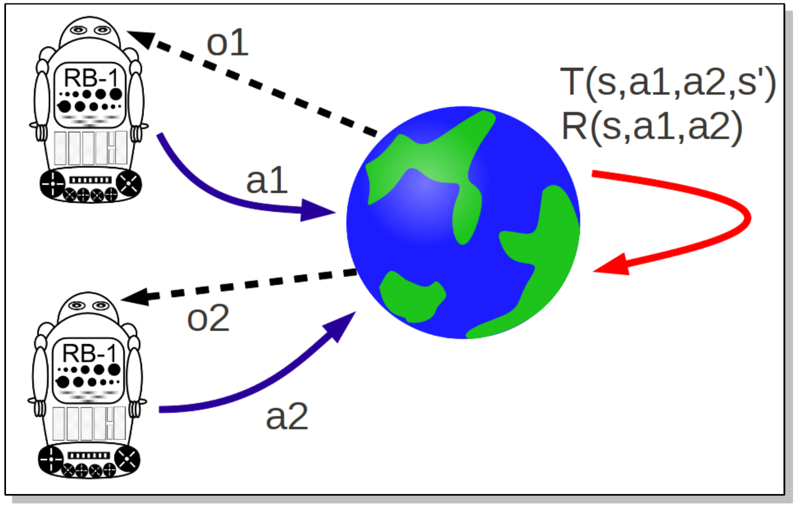


- Partial Observability



- Multiagent Systems: uncertainty about others

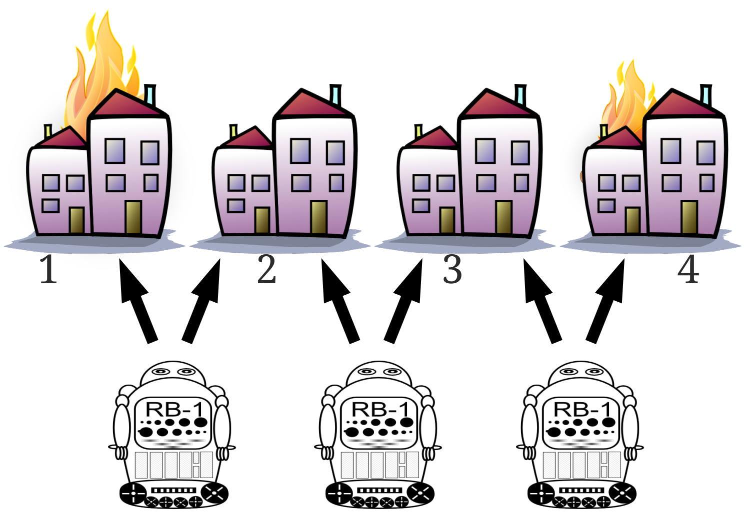
Formal Model

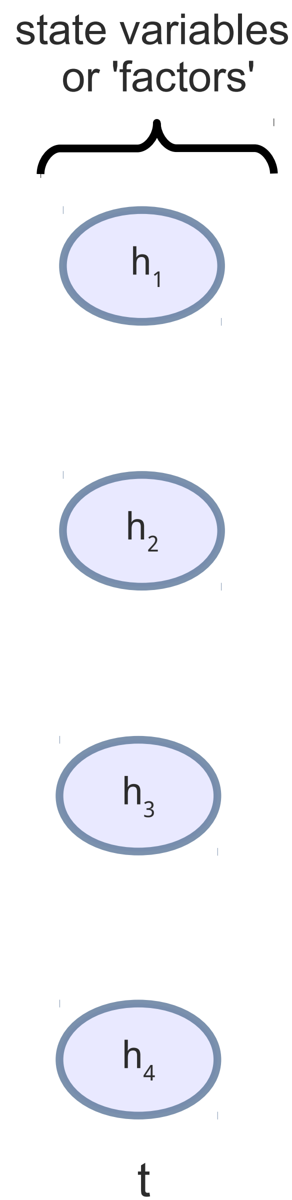


- A Dec-POMDP
  - $\langle S, A, P_T, O, P_O, R, h \rangle$
  - n agents
  - S – set of states
  - A – set of **joint** actions $\quad a = \langle a_1, a_2, \ldots, a_n \rangle$
  - P_T – transition function $\quad P(s' | s, a)$
  - O – set of **joint** observations $\quad o = \langle o_1, o_2, \ldots, o_n \rangle$
  - P_O – observation function $\quad P(o | a, s')$
  - R – reward function $\quad R(s,a) = E_{s'} R(s,a,s')$
  - h – horizon (finite)

Factored Dec-POMDPs

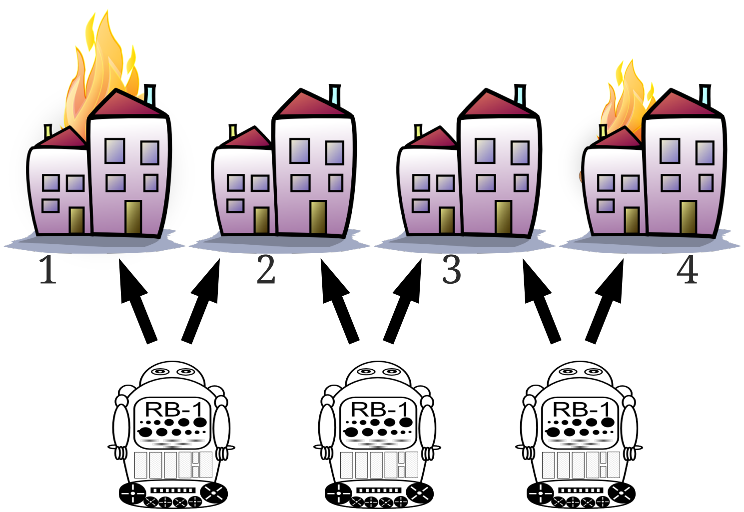
- E.g., 'Firefighting Graph'
  - H houses $h_1 \ldots h_H$
  - H-1 agents
    - actions: left or right
    - observations: flames or not

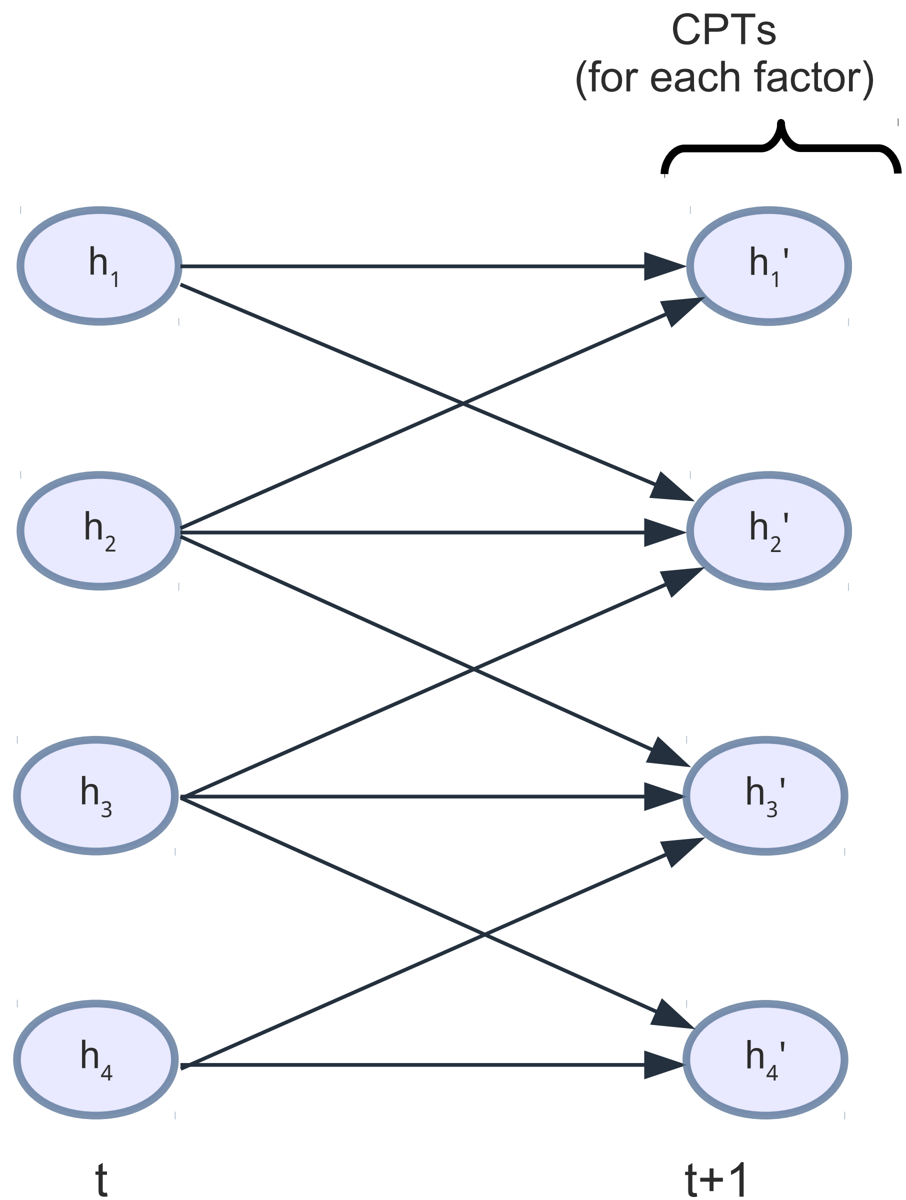




Factored Dec-POMDPs

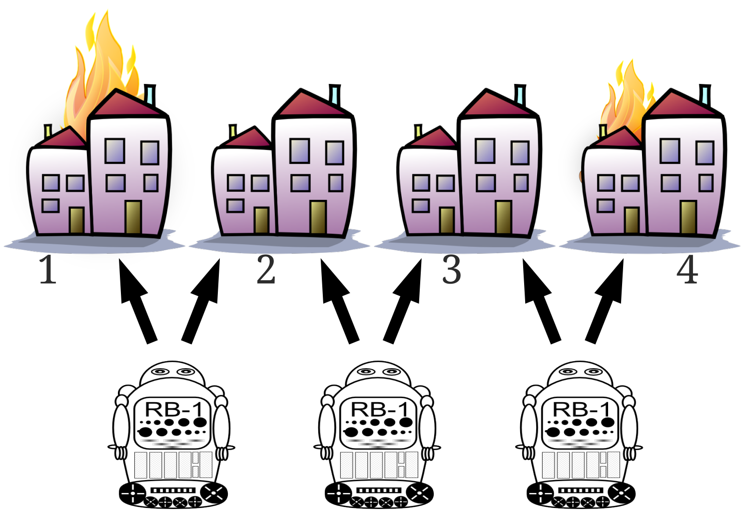
- E.g., 'Firefighting Graph'
  - H houses $h_1 \ldots h_H$
  - H-1 agents
    - actions: left or right
    - observations: flames or not

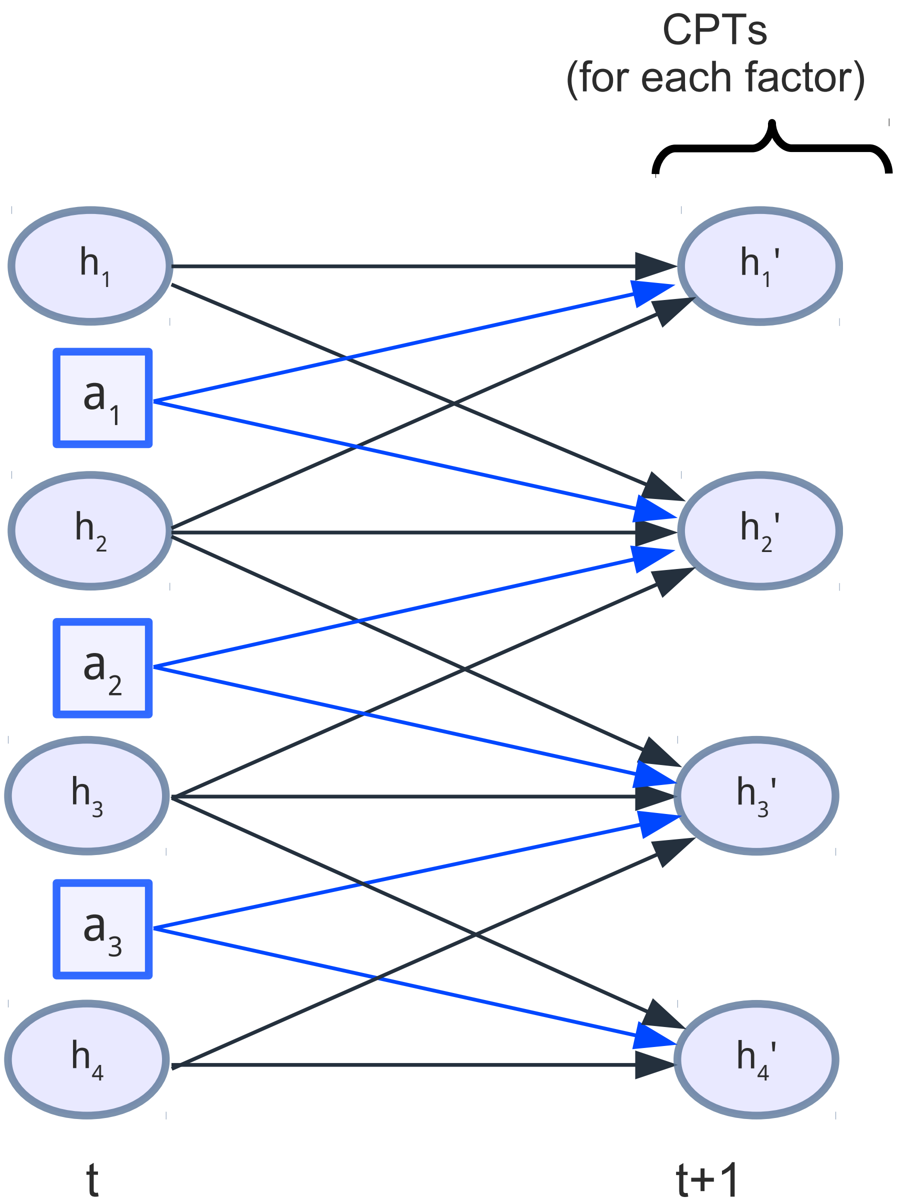




Factored Dec-POMDPs

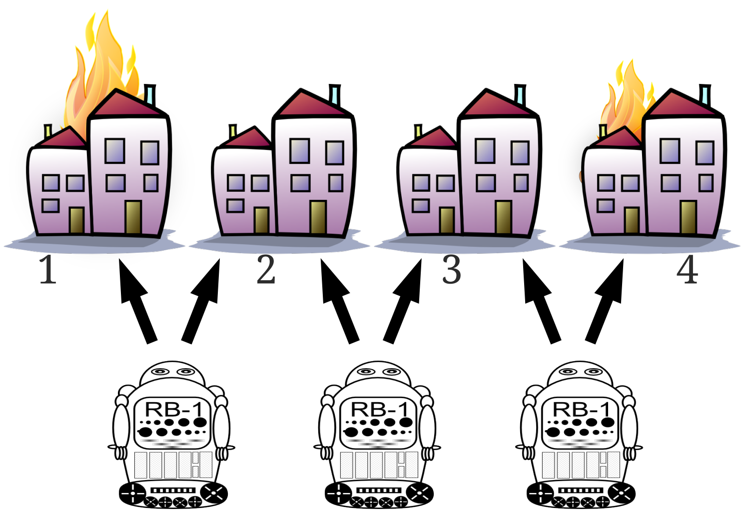
- E.g., 'Firefighting Graph'
  - H houses $h_1 \ldots h_H$
  - H-1 agents
    - actions: left or right
    - observations: flames or not

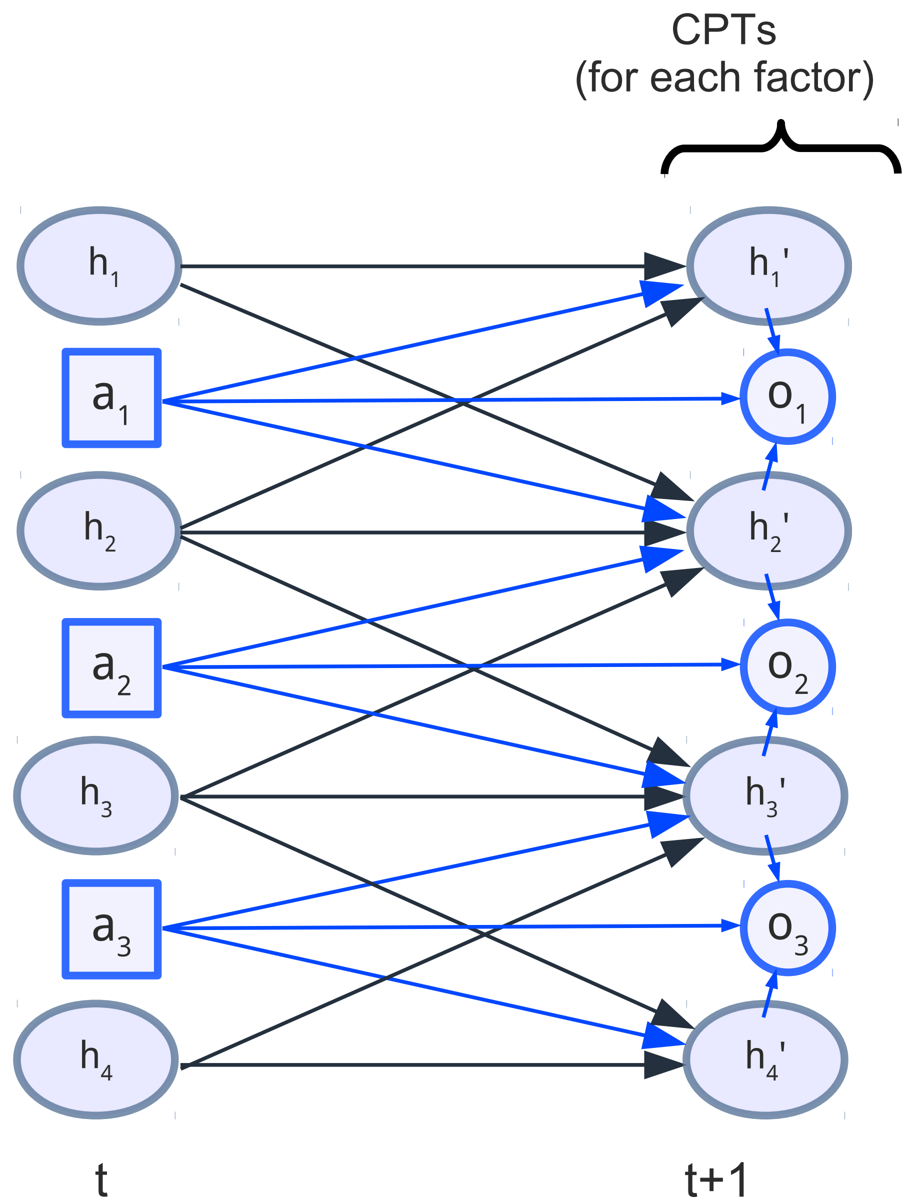




Factored Dec-POMDPs

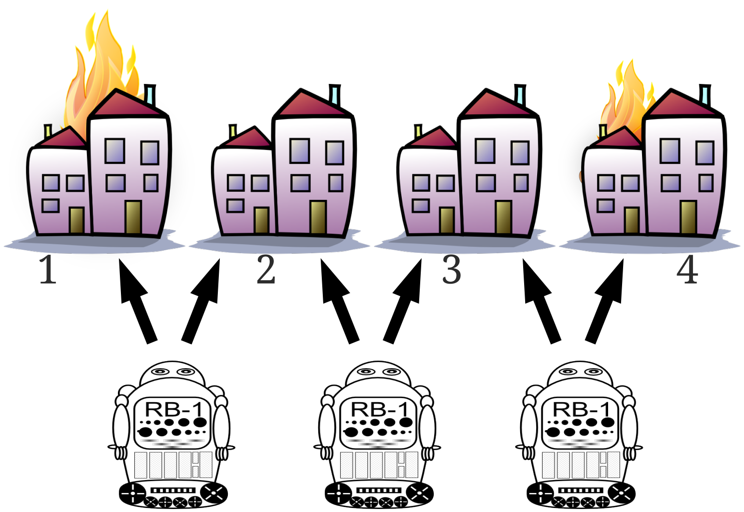
- E.g., 'Firefighting Graph'
  - H houses $h_1 \ldots h_H$
  - H-1 agents
    - actions: left or right
    - observations: flames or not

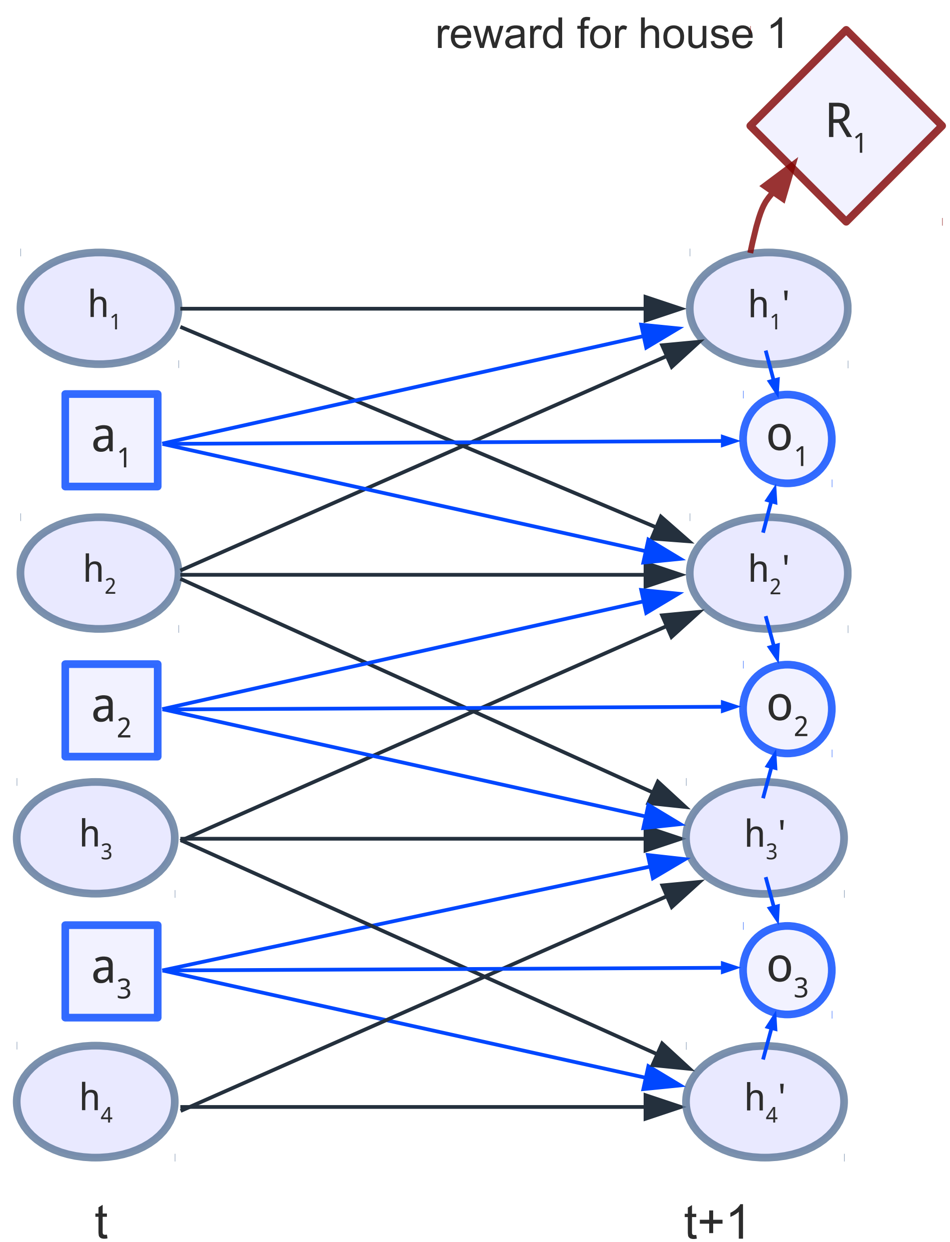




Factored Dec-POMDPs

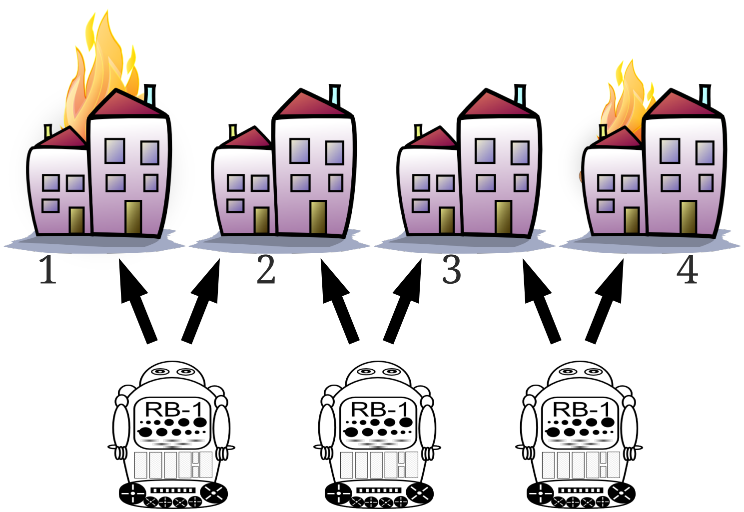
- E.g., 'Firefighting Graph'
  - H houses $h_1 \ldots h_H$
  - H-1 agents
    - actions: left or right
    - observations: flames or not

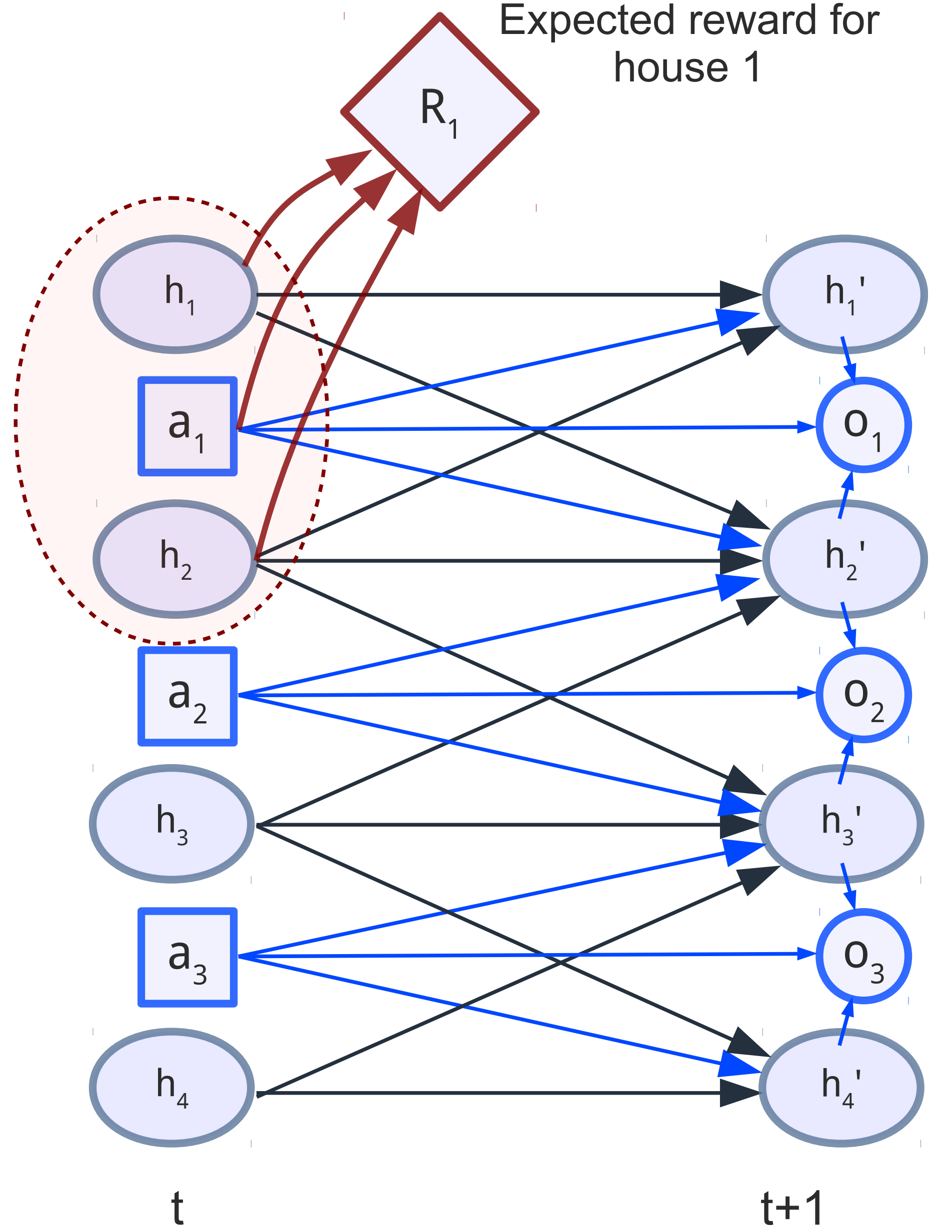




Factored Dec-POMDPs

- E.g., 'Firefighting Graph'
  - H houses $h_1 \ldots h_H$
  - H-1 agents
    - actions: left or right
    - observations: flames or not



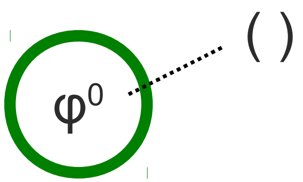


Background: FSPC – 1

- Goal: scalability w.r.t. #agents
- Forward-sweep policy computation (FSPC)
  - for each t=0,...,h-1
  - compute best joint decision rule δ^t
  - given past joint policy φ^t

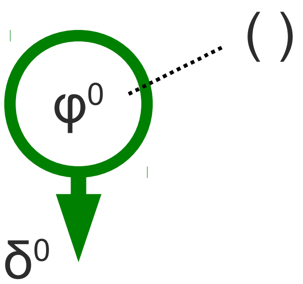
Background: FSPC – 1

- Goal: scalability w.r.t. #agents
- Forward-sweep policy computation (FSPC)
  - for each t=0,...,h-1
  - compute best joint decision rule δ^t
  - given past joint policy φ^t



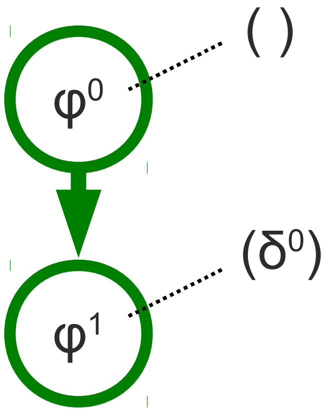
Background: FSPC – 1

- Goal: scalability w.r.t. #agents
- Forward-sweep policy computation (FSPC)
  - for each t=0,...,h-1
  - compute best joint decision rule δ^t
  - given past joint policy φ^t



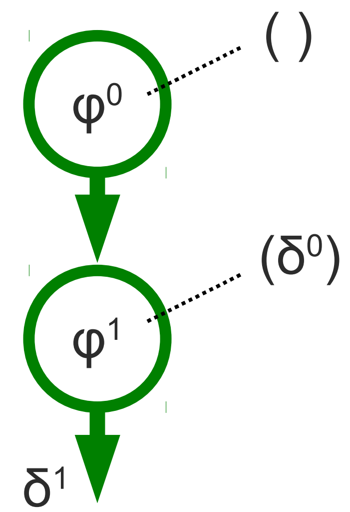
Background: FSPC – 1

- Goal: scalability w.r.t. #agents
- Forward-sweep policy computation (FSPC)
  - for each t=0,...,h-1
  - compute best joint decision rule δ^t
  - given past joint policy φ^t



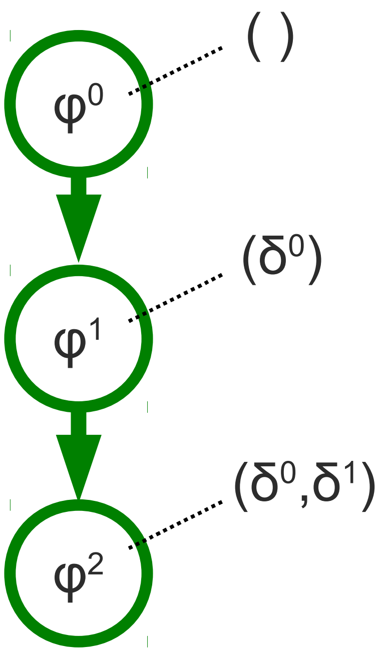
Background: FSPC – 1

- Goal: scalability w.r.t. #agents
- Forward-sweep policy computation (FSPC)
  - for each t=0,...,h-1
  - compute best joint decision rule δ^t
  - given past joint policy φ^t



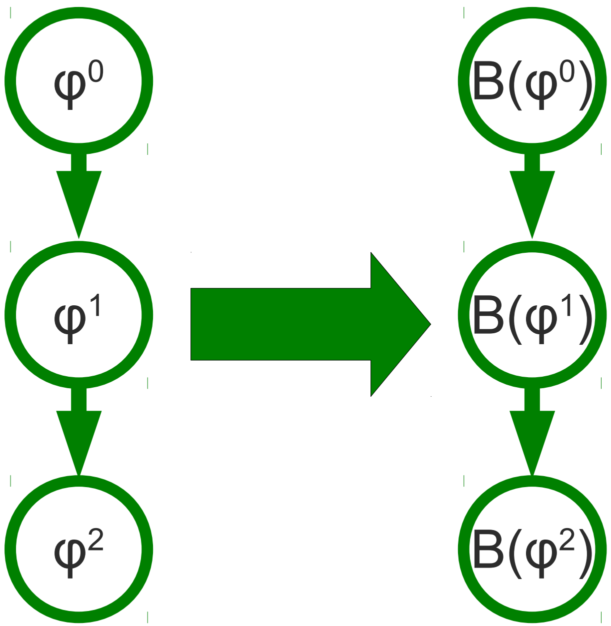
Background: FSPC – 1

- Goal: scalability w.r.t. #agents
- Forward-sweep policy computation (FSPC)
  - for each t=0,...,h-1
  - compute best joint decision rule δ^t
  - given past joint policy φ^t

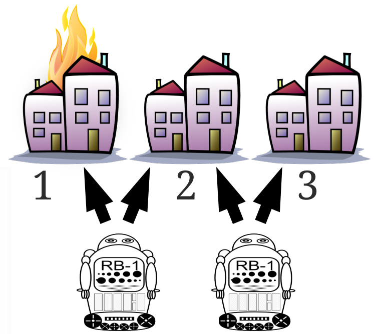


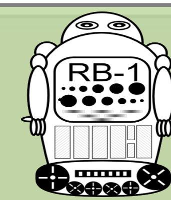
Background: FSPC – 2

- Problem at stage t
  - select $\delta = <\delta_1,...,\delta_n>$
  - δ_i: histories → actions $\delta_i^t(\vec{\theta}_i^t) = a_i^t$
- As **collaborative Bayesian games (CBGs)**
  - agents, actions
  - types θ_i ↔ histories
  - probabilities: $P(\theta)$
  - payoffs: $Q(\theta,a)$

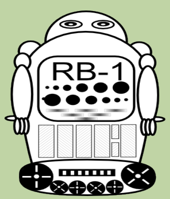


Background: CBGs



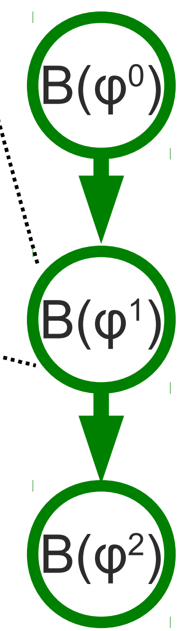


ag. 2



agent 1

		Flames		No Flames	
		H2	H3	H2	H3
Flames	H1	+2	-1	...	...
	H2	...	...	...	...
No Flames	H1	...	...	...	...
	H2	...	...	...	...



- As **collaborative Bayesian games (CBGs)**
  - agents, actions
  - types $\theta_i \leftrightarrow$ histories
  - probabilities: $P(\theta)$
  - payoffs: $Q(\theta,a)$

This paper: Factored FSPC

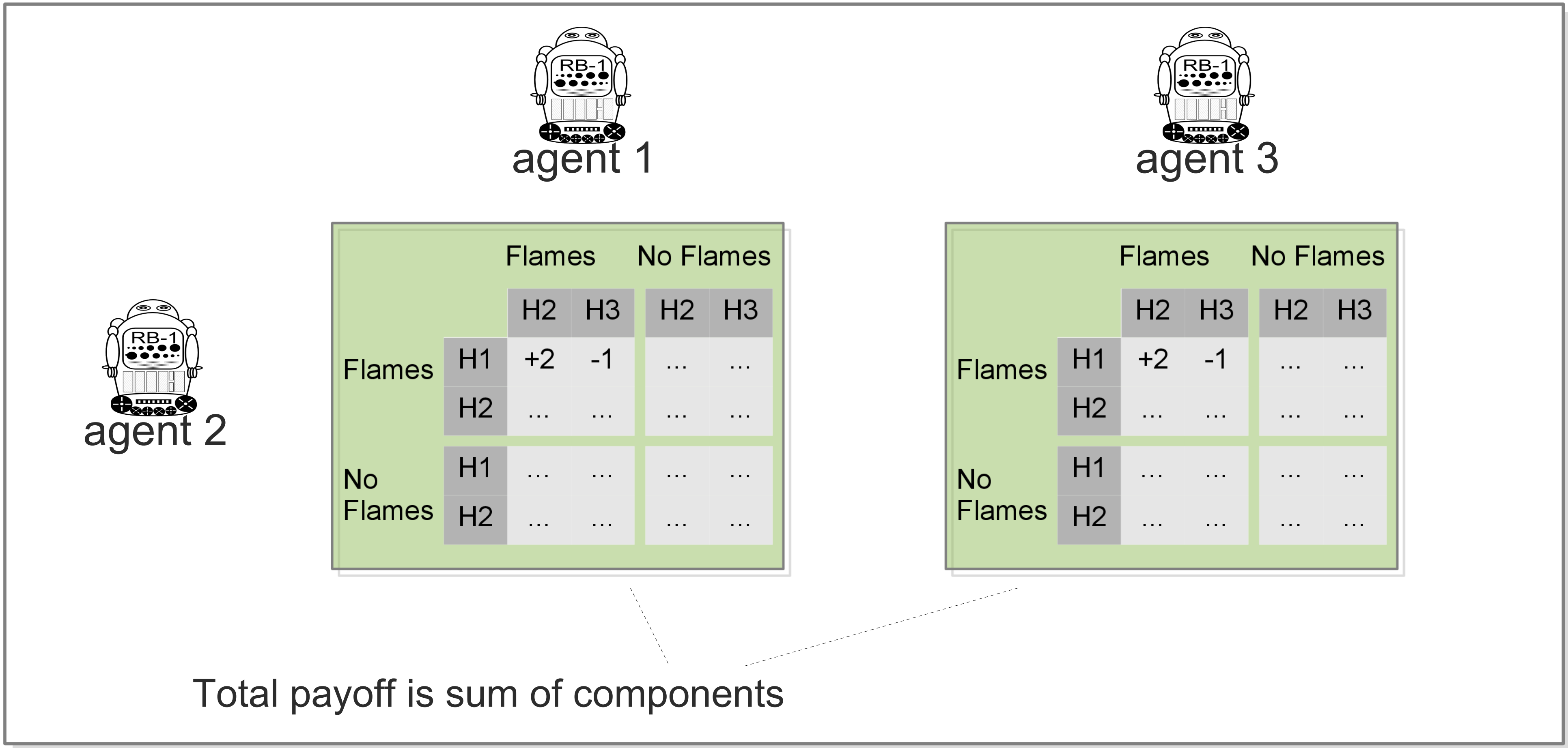
Factored FSPC – Basic idea:

exploit independence between agents in FSPC

- if value function factored $Q(x, \vec{\theta}, a) = \sum_e Q_e(x_e, \vec{\theta}_e, a_e)$

  → replace CBGs with
  Collaborative **graphical** Bayesian games (CGBGs)

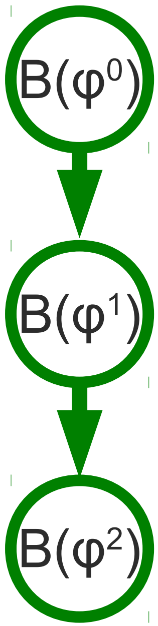
This paper: Factored FSPC



→ replace CBGs with

Collaborative **graphical** Bayesian games (CGBGs)

Factored FSPC for Many Agents



Improving scalability w.r.t. the number of agents...

1) Approximate structure of *Q**

- **Predetermined scope structure**

2) Compute CGBG payoff functions

- **Transfer Planning (TP)**

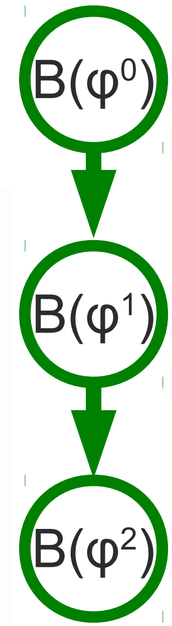
3) Inference techniques to construct CGBGs

- Extension of factored frontier [Murphy&Weiss UAI 2001]

4) Efficient solutions of CGBGs

- Max-plus to ATI-FG [OWS UAI 2012]

Factored FSPC for Many Agents



Improving scalability w.r.t. the number of agents...

1) Approximate structure of *Q**

- **Predetermined scope structure**

2) Compute CGBG payoff functions

- **Transfer Planning (TP)**

Rest of this talk

3) Inference techniques to construct CGBGs

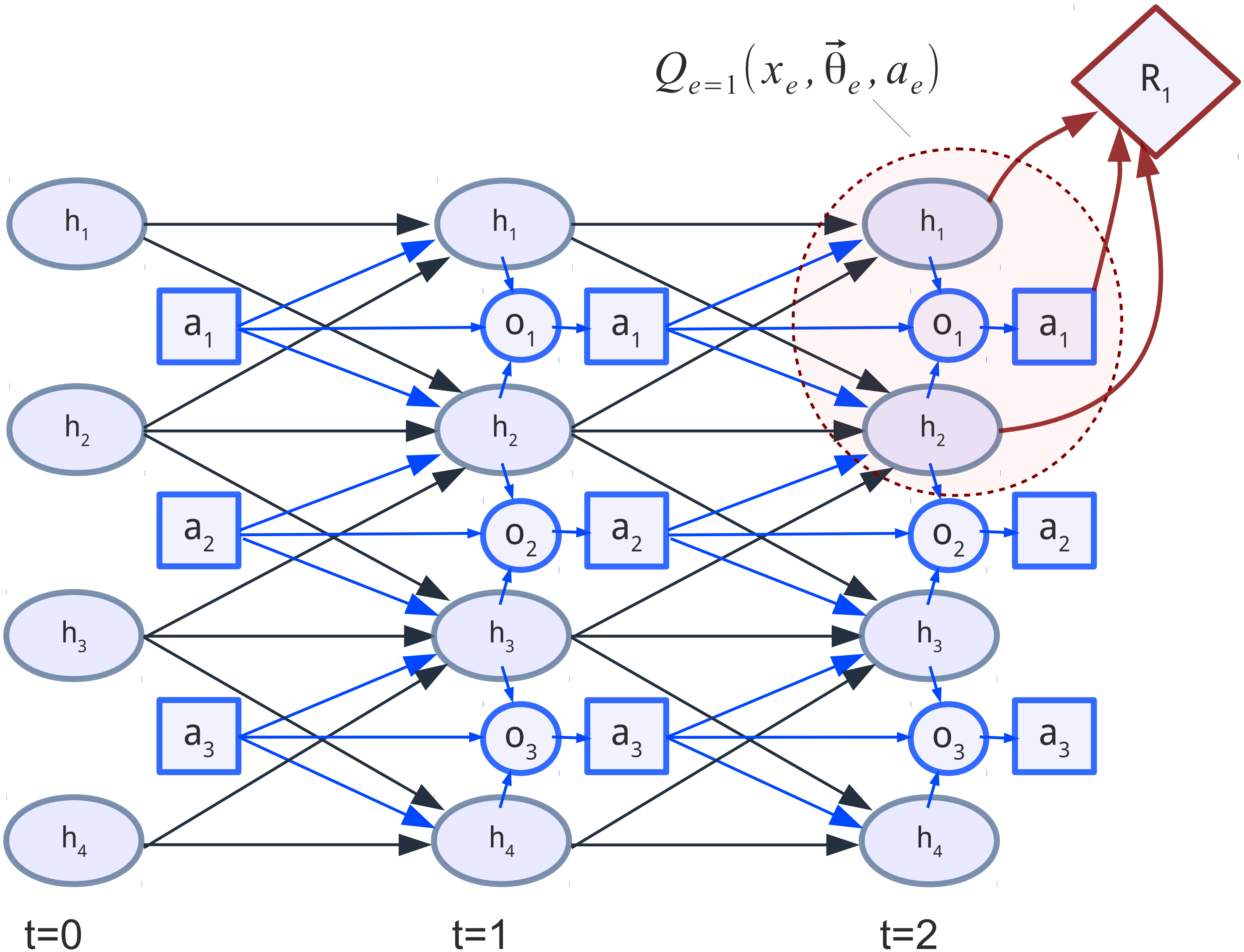
- Extension of factored frontier [Murphy&Weiss UAI 2001]

4) Efficient solutions of CGBGs

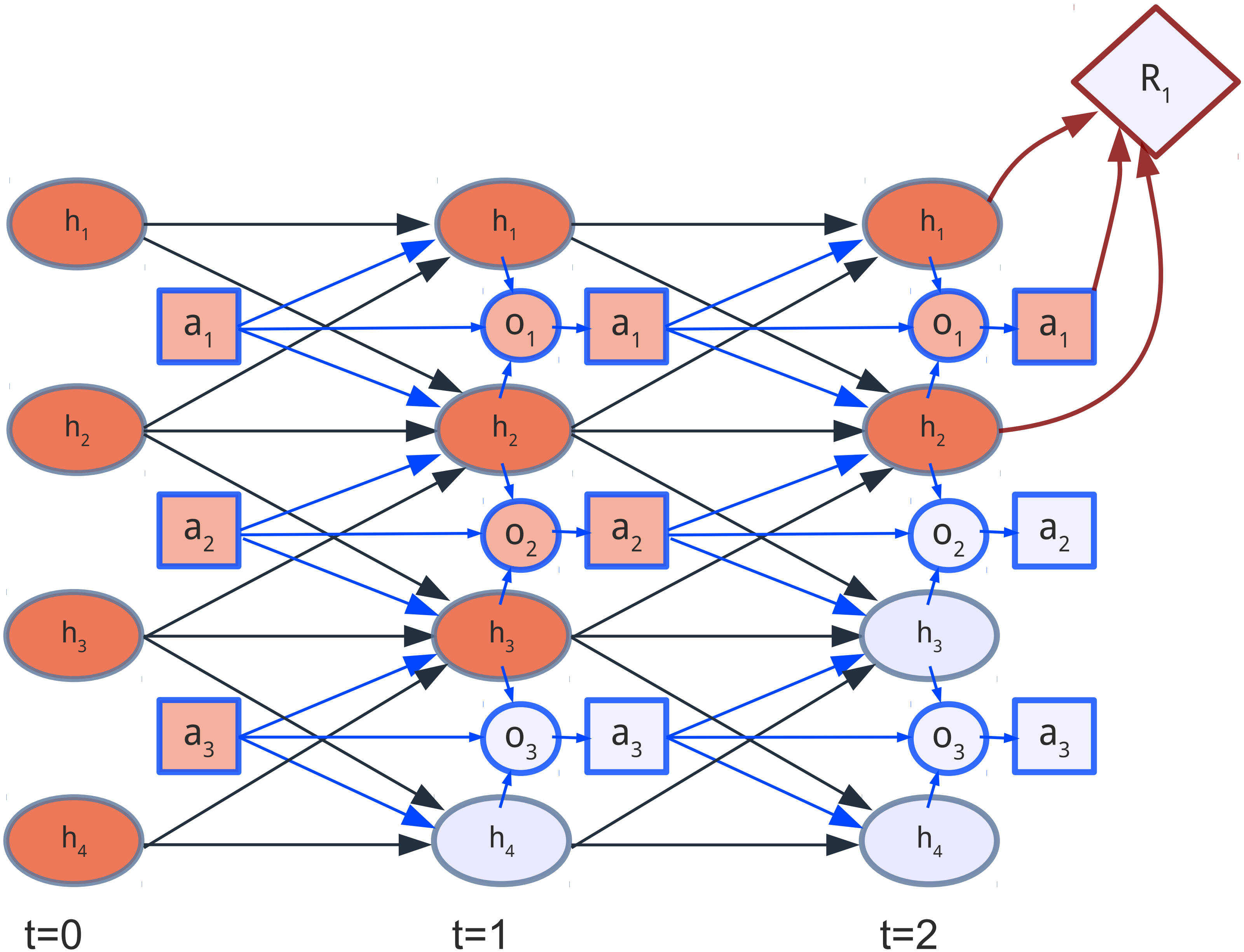
- Max-plus to ATI-FG [OWS UAI 2012]

see paper

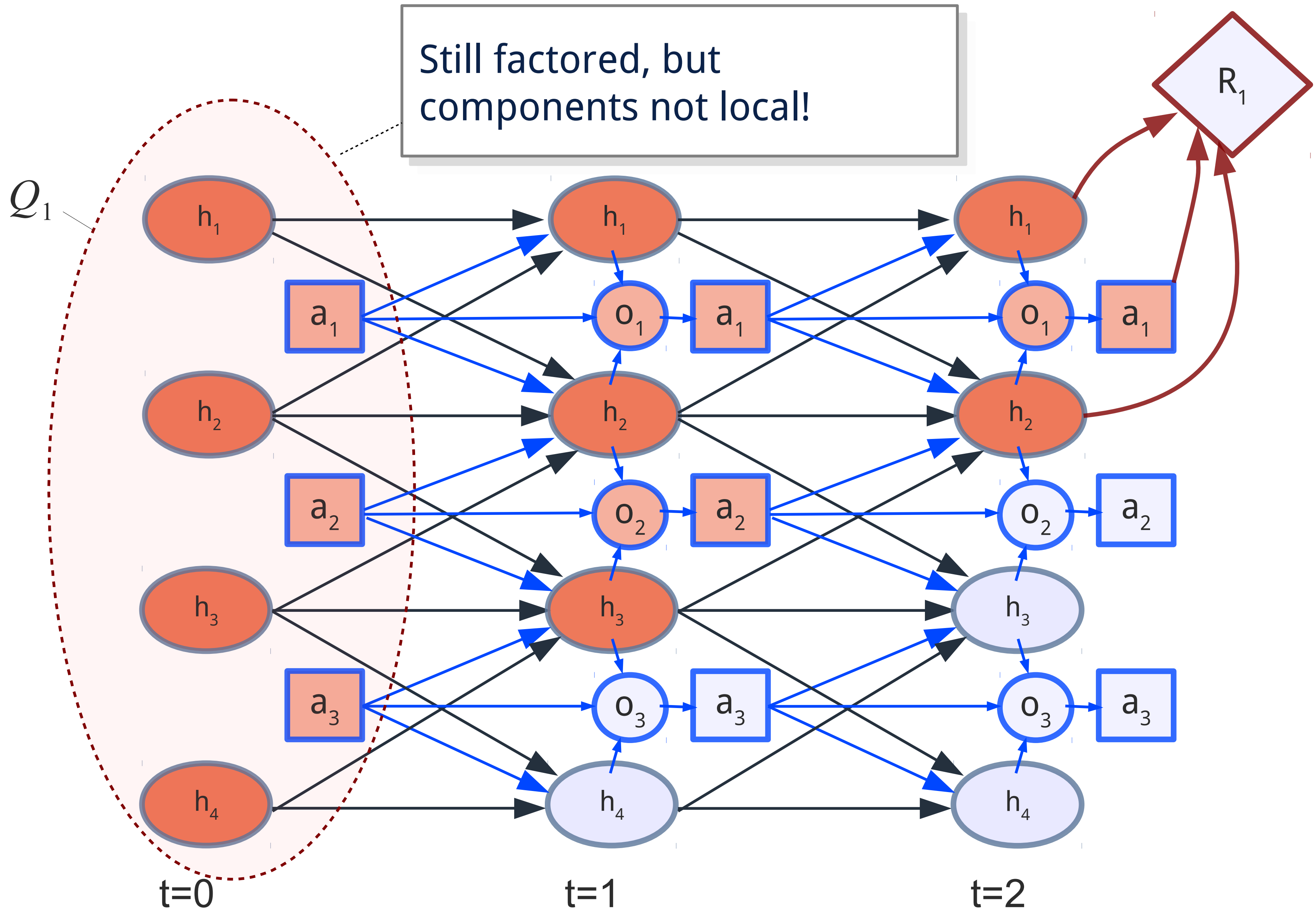
Accumulation of Reward over Time



Accumulation of Reward over Time

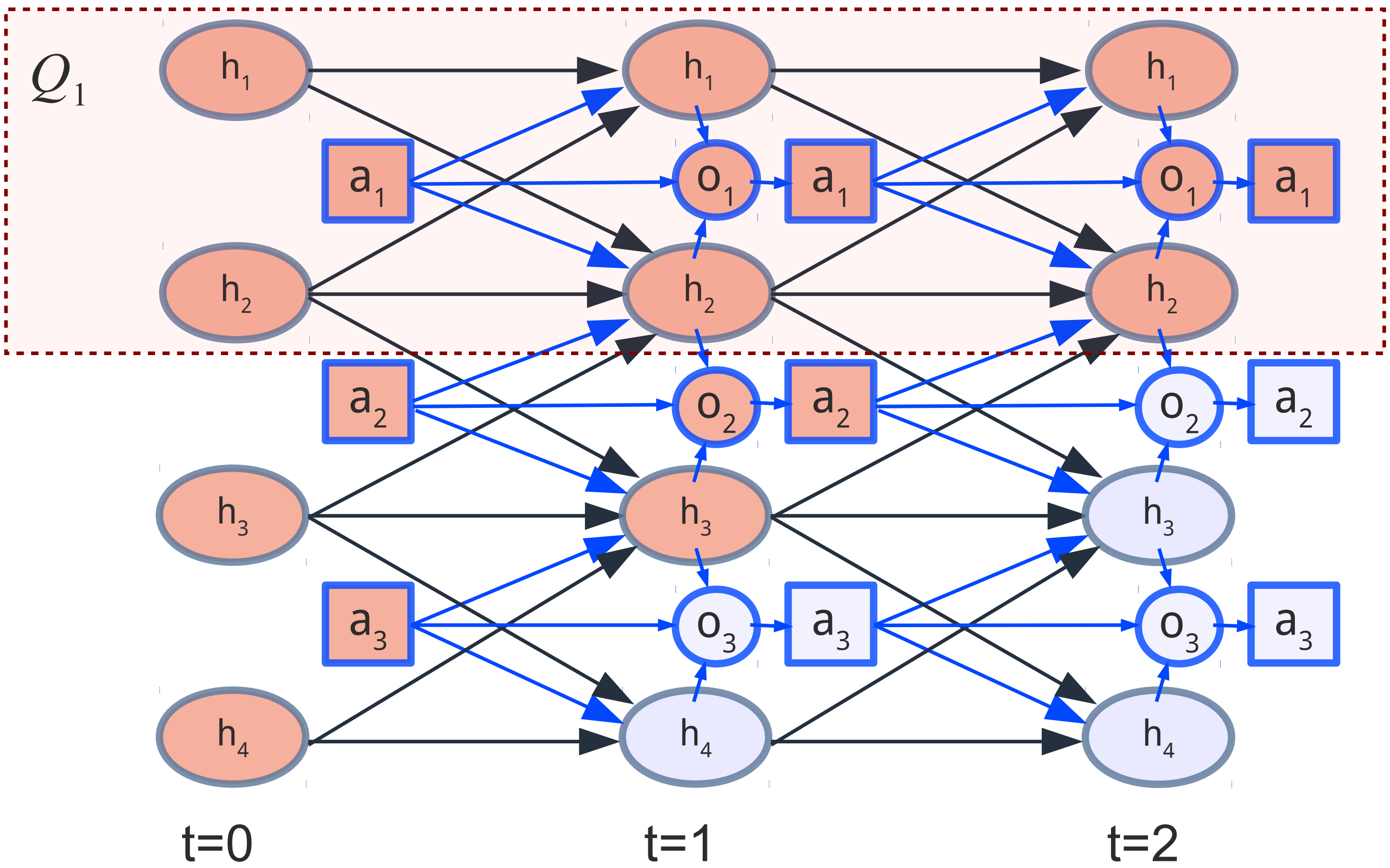


Accumulation of Reward over Time



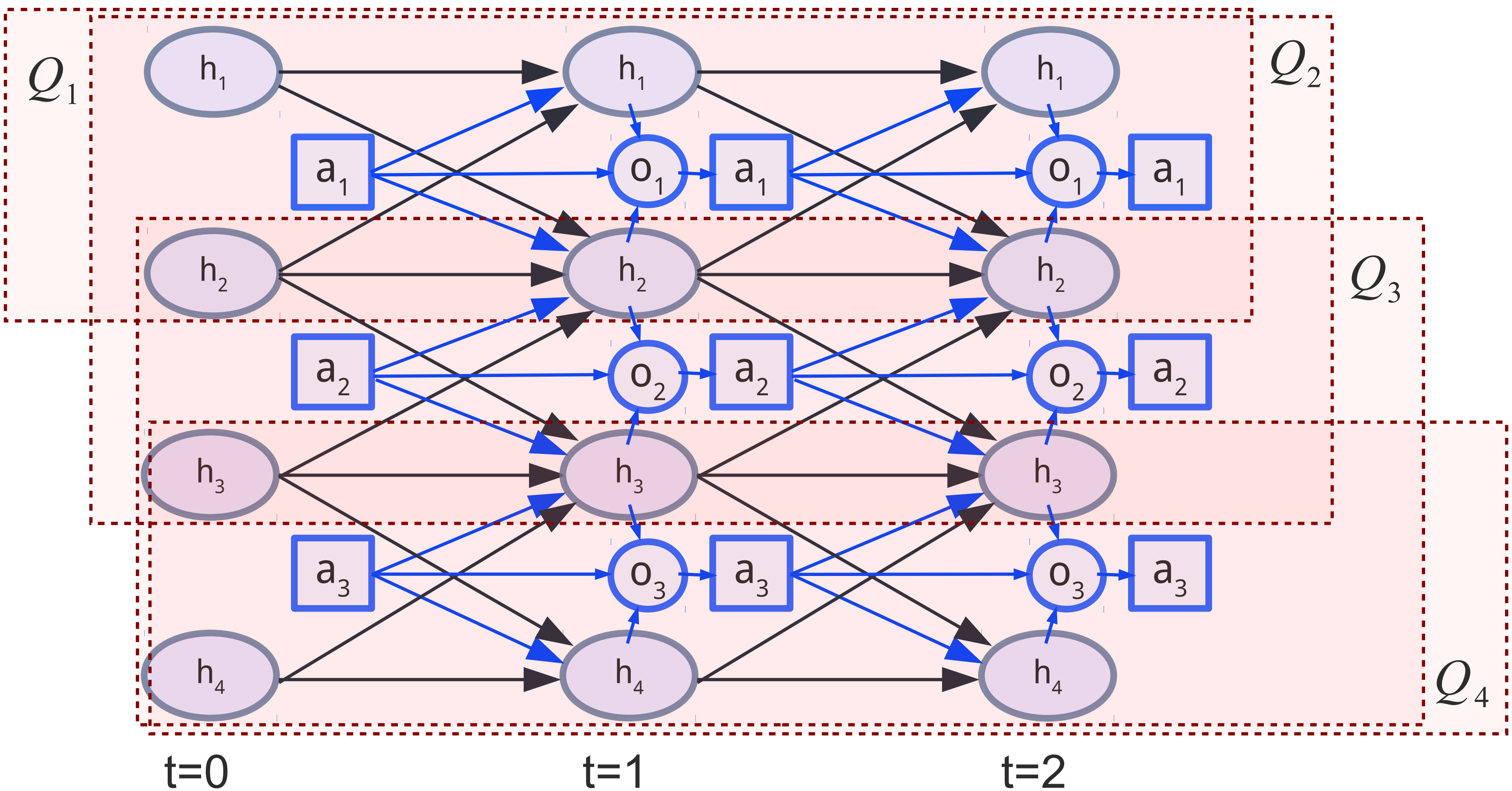
1) Predetermined Scope Structure

- Use smaller scopes



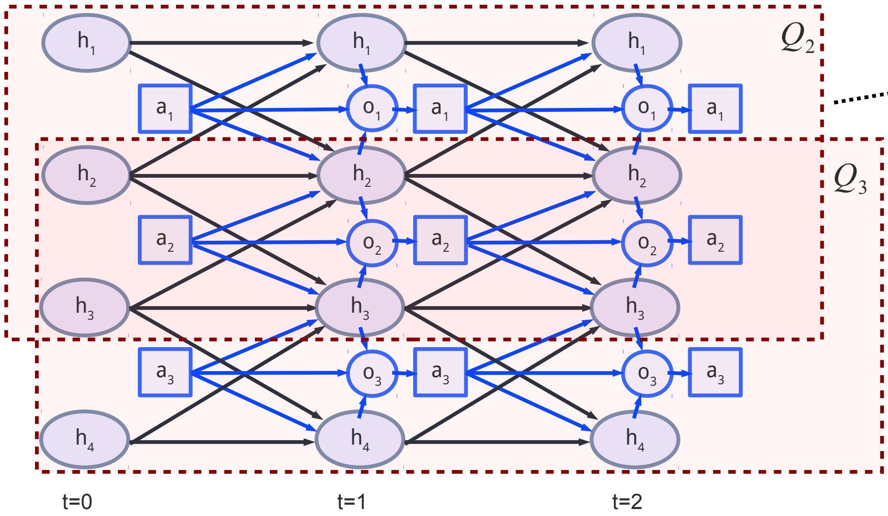
1) Predetermined Scope Structure

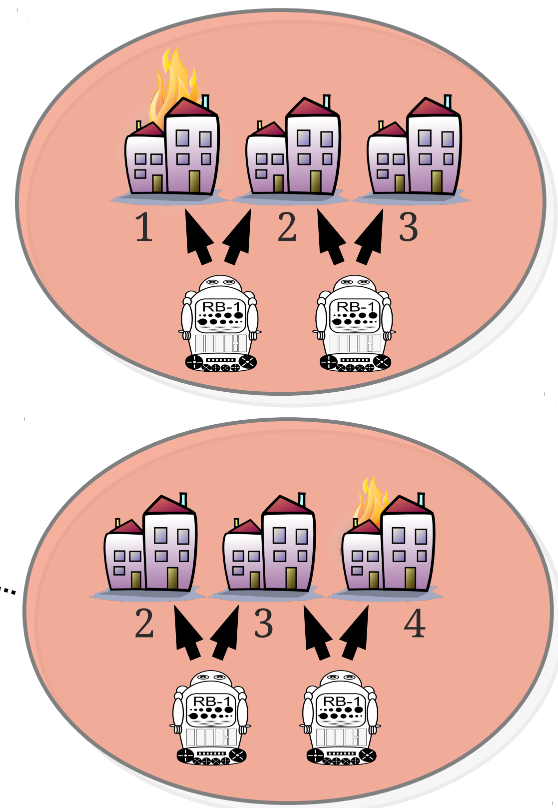
- Use smaller scopes



2) Transfer Planning

- Define a **source problem** for each component
  - involving fewer agents

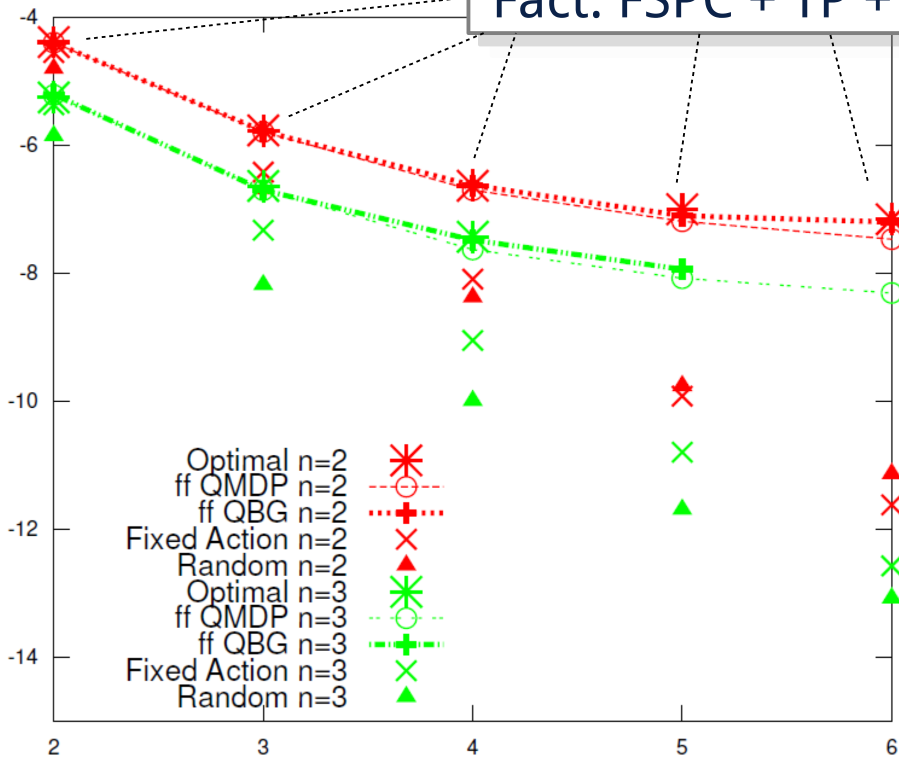




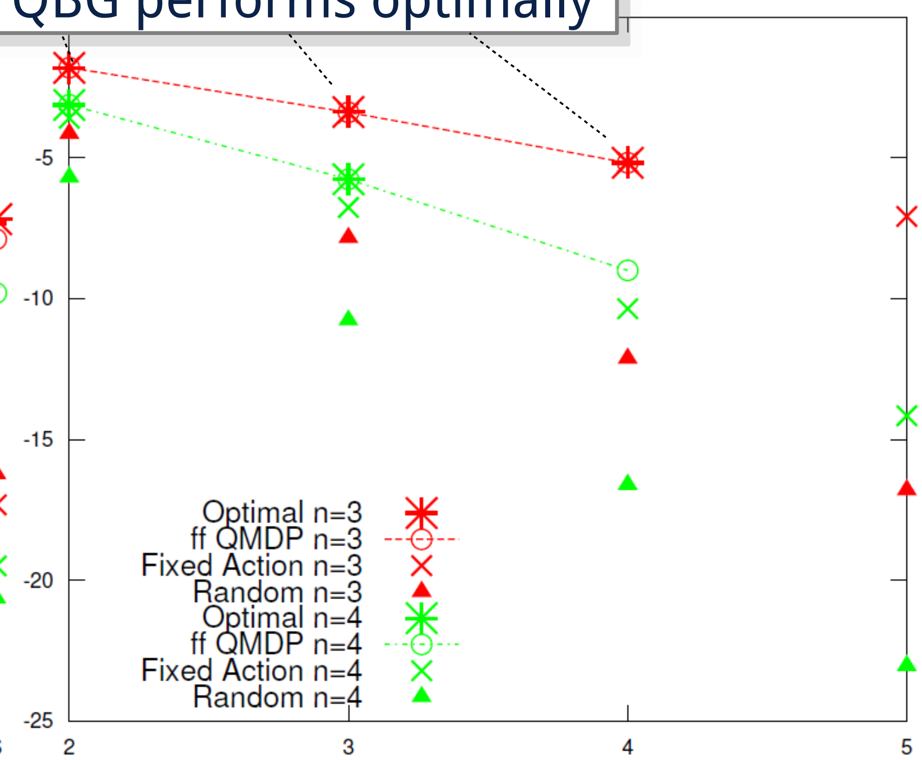
- Solve those exactly or approximately (QMDP, etc.)

Results – Compared to Optimal

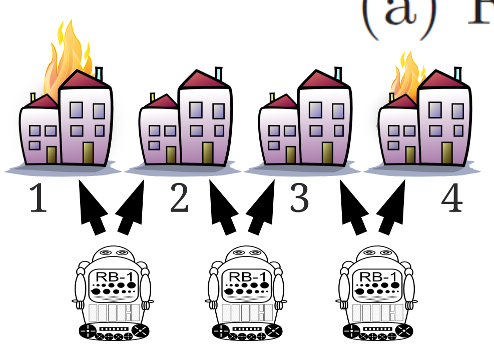
Fact. FSPC + TP + QBG performs optimally

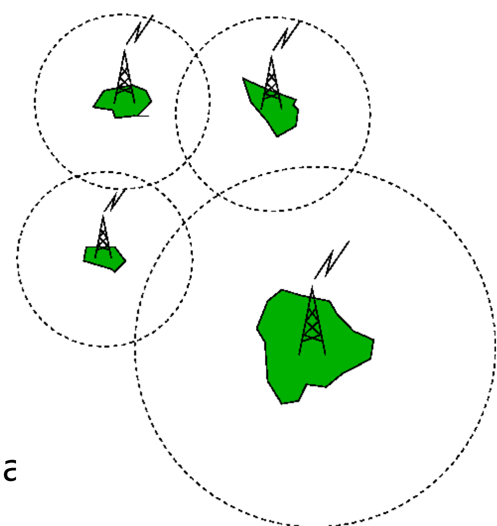


(a) FFG.



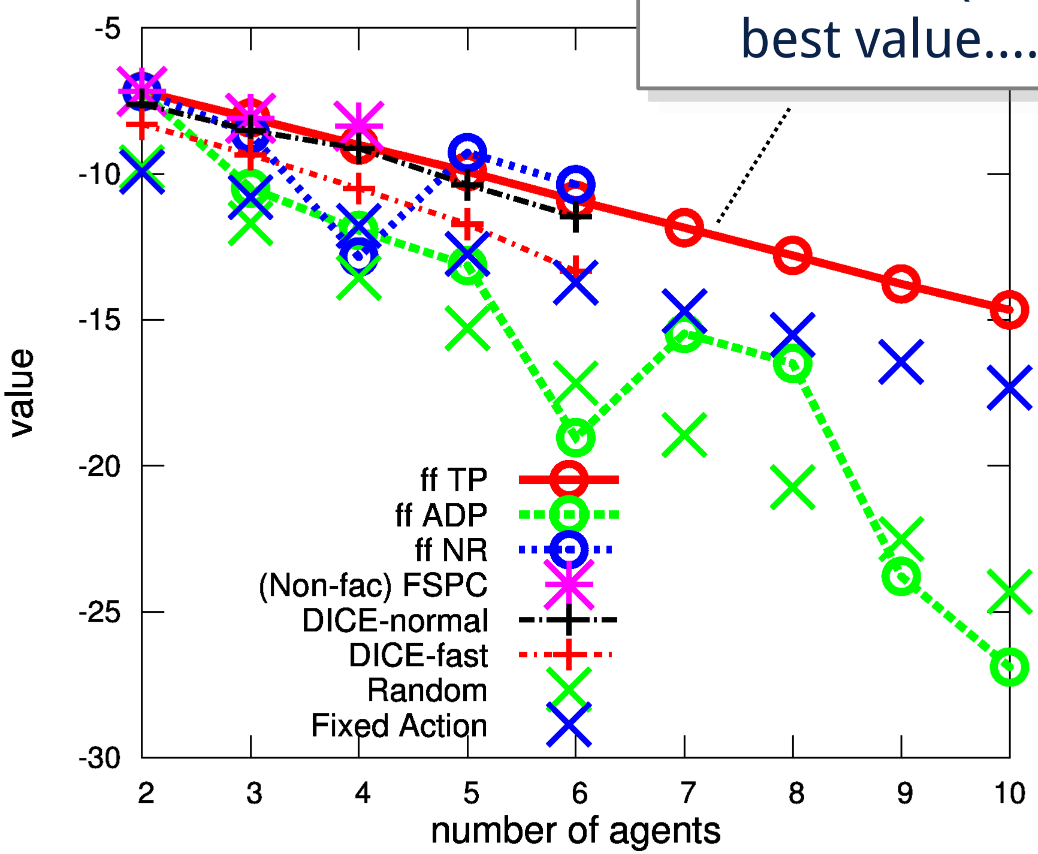
(b) Aloha.



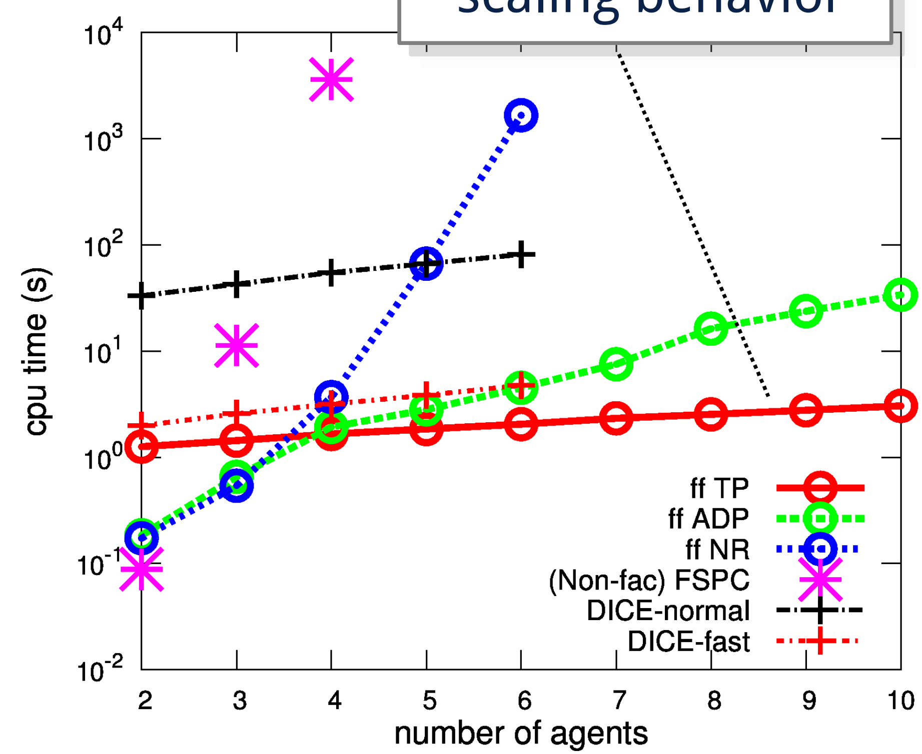


Results – vs. Approximate

TP achieves (near-) best value....

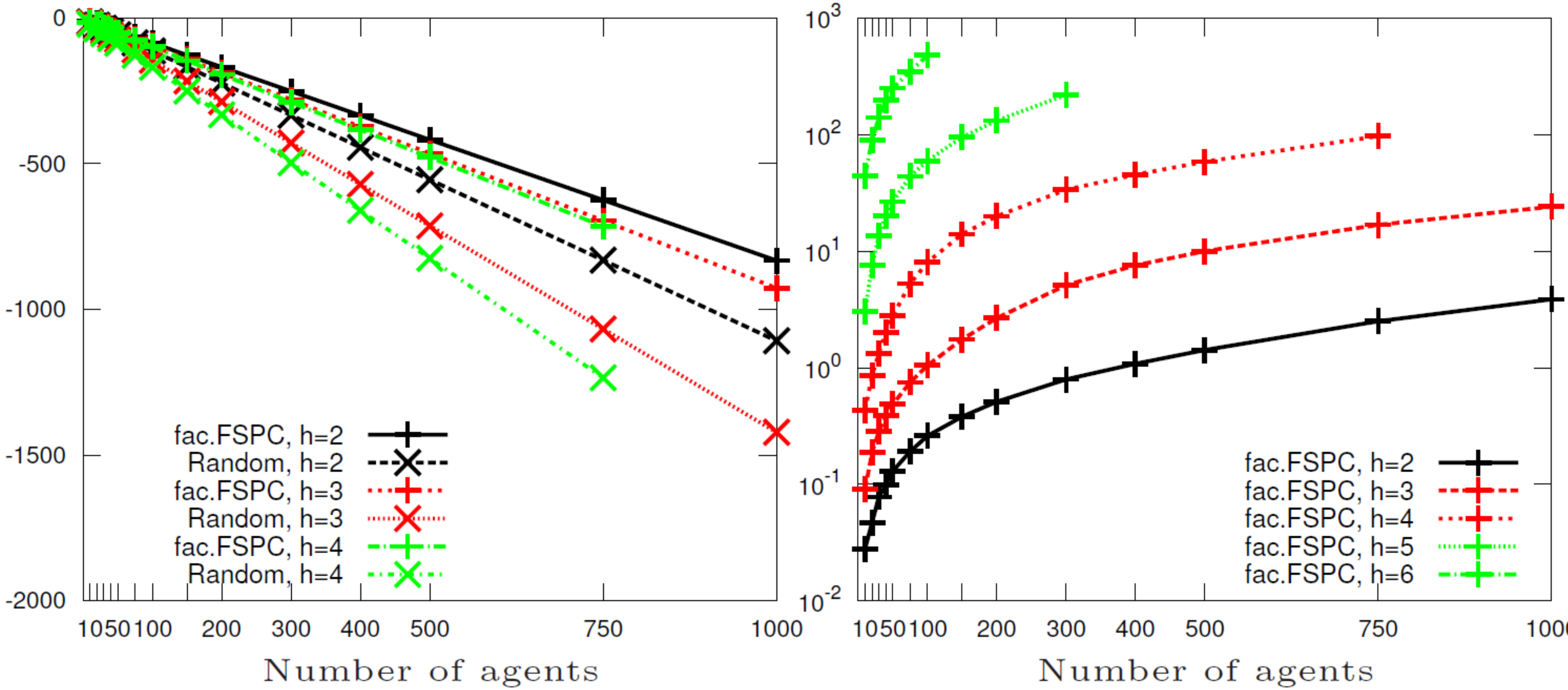


...and superior scaling behavior



- FFSCP: TP vs ADP, NR
- Non-factored FSPC, DICE [Oliehoek et al. 2008 Informatica]

Results – Many Agents



(a) Value. $h = 2 \ldots 4$. (b) Runtime. $h = 2 \ldots 6$.

Conclusions

- Factored FSPC with transfer planning:

  approximates factored Dec-POMDPs with multiple abstractions involving subsets of agents

- Unprecedented scalability for this class
  - results up to 1000 agents

- Future work:
  - scale to higher horizon
  - understand such abstractions
    - empirically verified near-optimal quality
    - formal understanding influence-based abstraction [AAAI' 12]

References

- R. Emery-Montemerlo, G. Gordon, J. Schneider, and S. Thrun. Approximate solutions for partially observable stochastic games with common payoffs. In AAMAS, 2004.
- K. P. Murphy and Y. Weiss. The factored frontier algorithm for approximate inference in DBNs. In UAI, 2001.
- F. A. Oliehoek, J. F. Kooi, and N. Vlassis. The cross-entropy method for policy search in decentralized POMDPs. Informatica, 32:341–357, 2008
- F. A. Oliehoek, M. T. J. Spaan, S. Whiteson, and N. Vlassis. Exploiting locality of interaction in factored Dec-POMDPs. In AAMAS, 2008.
- F. A. Oliehoek, S. Witwicki, and L. P. Kaelbling. Influence-Based Abstraction for Multiagent Systems. In AAAI, 2012.
- F. A. Oliehoek, S. Whiteson, and M. T. J. Spaan. Exploiting structure in cooperative Bayesian games. In UAI, 2012.
- D. Szer, F. Charpillet, and S. Zilberstein. MAA*: A heuristic search algorithm for solving decentralized POMDPs. In UAI, 2005.