Exploiting Local Interactions

Francisco Melo (joint work with M. Spaan)

Planning, Execution and Learning

Manuela Veloso and Reid Simmons

Carnegie Mellon University

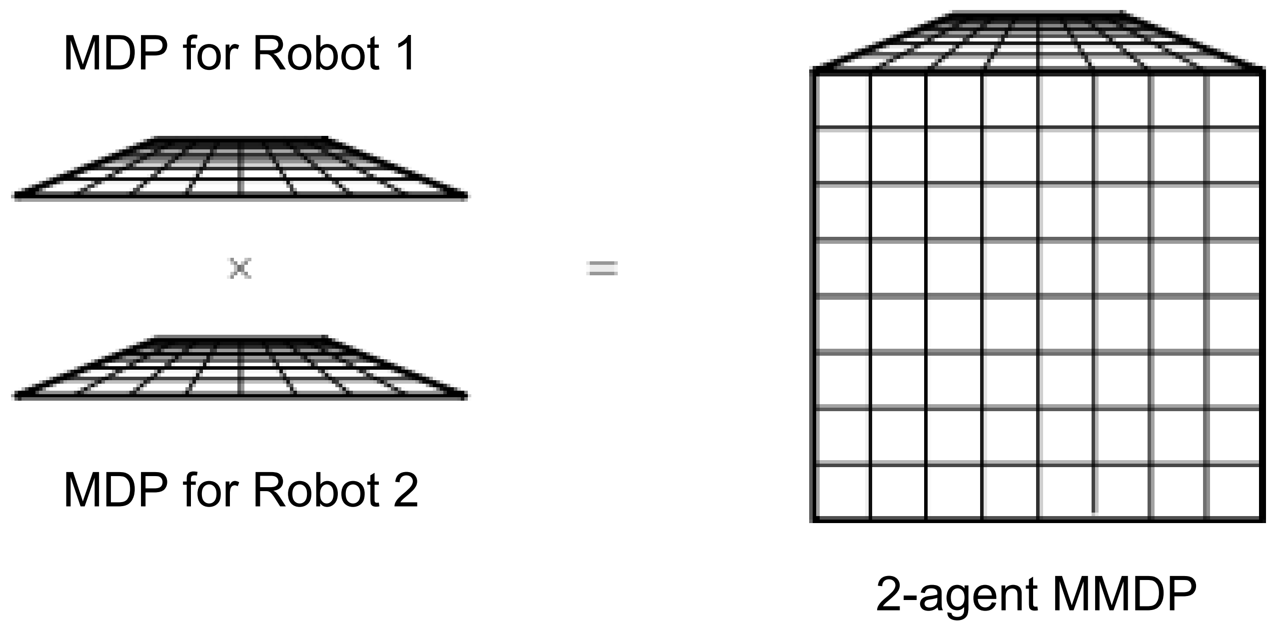
Fall 2008

Outline

- **MMDPs vs. Dec-POMDPs**
- Simplified Models
- Decoupling Interaction
- Approximate Solutions

MMDPs

- Multiagent MDPs: “Big” MDPs



- State/action spaces grow exponentially

MMDPs (cont.)

- Each robot has full state observability =

  Each robot knows its state and that of all other agents

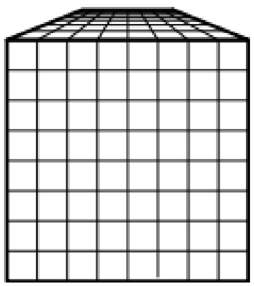
- Each robot has full action observability =

  Each robot knows its action and that of all other agents

Requires continuous communication and perfect sensing!

Dec-POMDPs

- Dec-POMDPs alleviate full observability assumptions



MMDP

+

Partial Observability

- Decisions based on incomplete world perception

Dec-POMDPs (cont.)

- Each robot has partial state observability =

  Each robot can only perceive incomplete local state information

- Each robot has partial action observability =

  Each robot knows only its own actions

Exact solution methods computationally intractable!

MMDPs vs. Dec-POMDPs

MMDPs	Dec-POMDPs
Solvable in reasonable amount of time	Not solvable in any reasonable amount of time
Irrealistic models of multi-robot interaction	Reasonable model of multi-robot interaction

Outline

- MMDPs vs. Dec-POMDPs
- **Simplified Models**
- Decoupling Interaction
- Approximate Solutions

Simplified Models I

- Dec-MDP
  - State is jointly fully observable
  - With “full” communication = MMDP

- However...
  - Complexity similar to Dec-POMDP (no comm)

Simplified Models II

- Dec-POMDP + Communication
  - With “full” communication = (M)POMDP
  - Different delays in communication can be used to obtain successively better approximations[1]

[1]F. Oliehoek, M. Spaan, N. Vlassis. Optimal and Approximate Q-value Functions for Dec-POMDPs. JAIR (32), pp. 289-353, 2008.

Simplified Models III

- Transition/Observation Independent Dec-MDPs
  - Agents coupled only by reward/value function
  - Breaks the problem into N “coupled” MDPs[2] – see last class

[2] R. Becker, S. Zilberstein, V. Lesser, C. Goldman. Solving Transition Independent Dec-MDPs. JAIR (22), pp. 423–455, 2004.

Outline

- MMDPs vs. Dec-POMDPs
- Simplified Models
- **Decoupling Interaction**
- Approximate Solutions

Complexity Issues

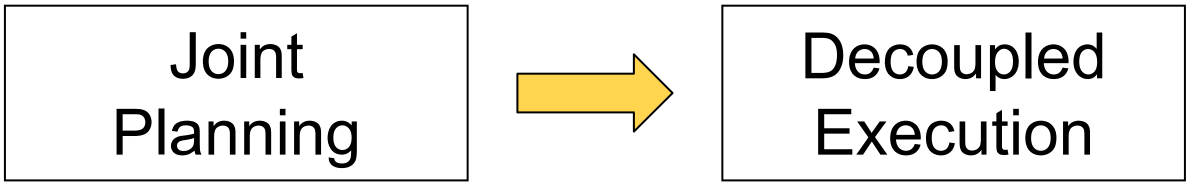
- Complexity comes from:
  - Partial observability
  - Number of agents

- Why exponential growth?
  - Coupling between agents (policy-wise)

Coupling comes from
joint reward functions

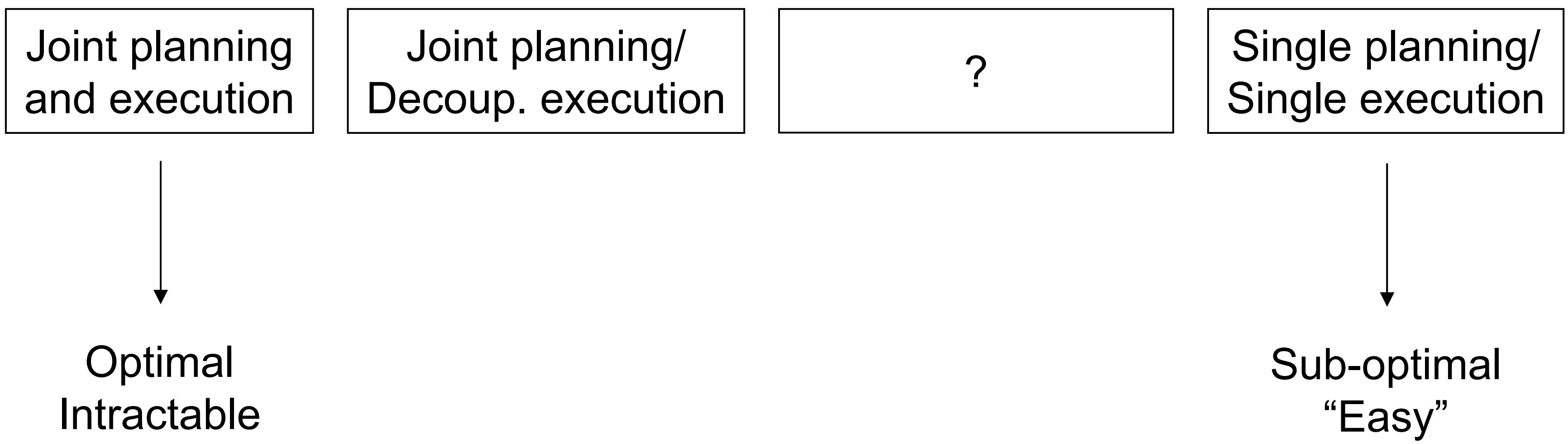
Decoupling Interaction I

- Decoupling execution (last class)
  - Alleviates requirements for continuous communication



- However...
  - Planning is still in joint space

Overview

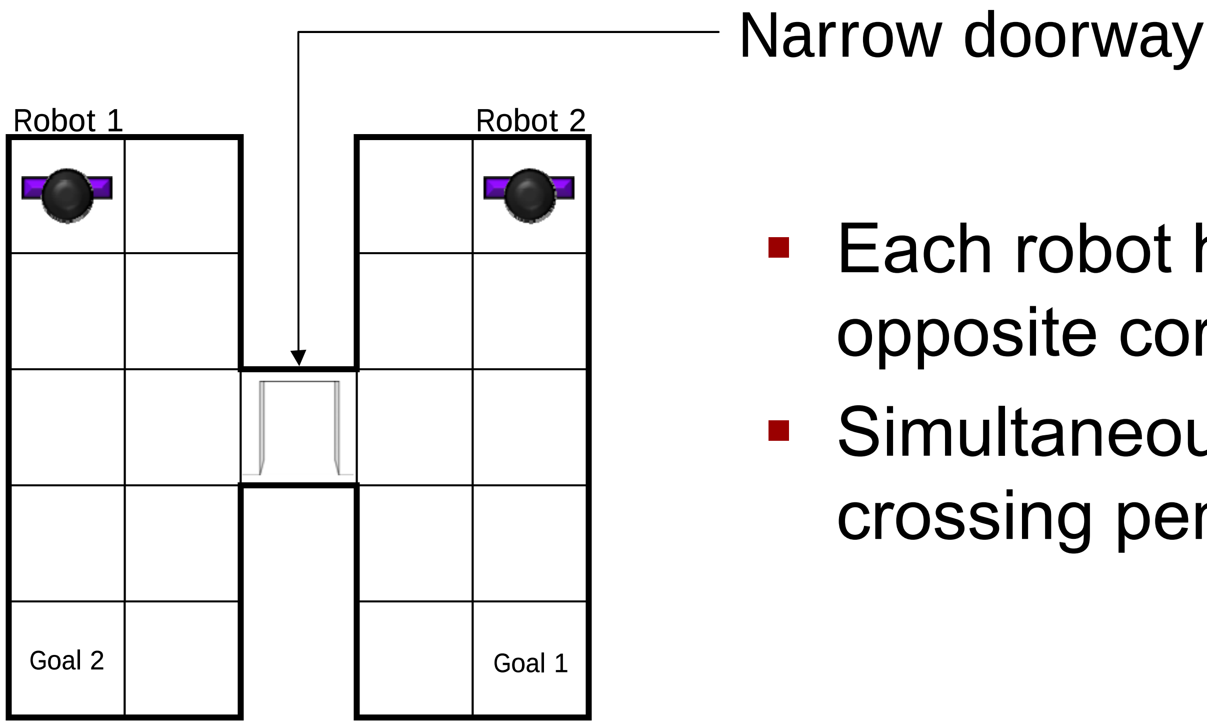

Joint planning and execution
Joint planning/ Decoup. execution
?
Single planning/ Single execution
Optimal
Intractable
Sub-optimal
“Easy”

Decoupling Interaction II

- Simplify policy dependencies
- Consider single agent policies
- “Add” interactions when needed

Example: The H World

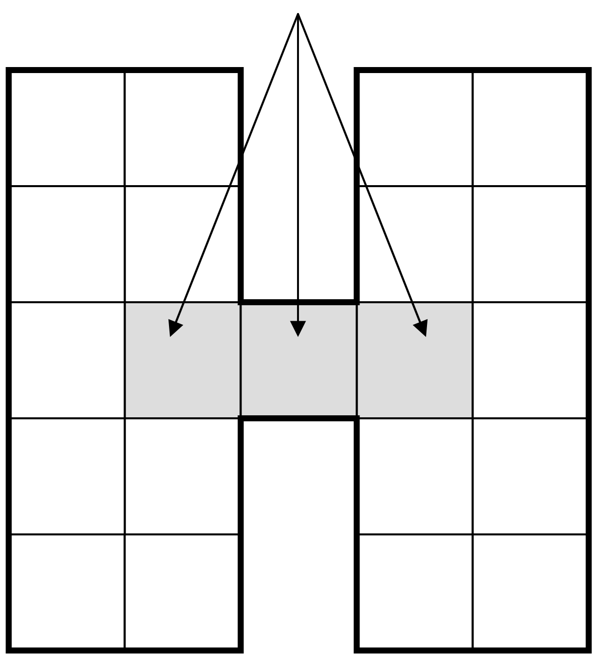
- Navigation example:



- Each robot has to reach opposite corner;
- Simultaneous doorway crossing penalized;

Example: The H World (cont.)

- Robots can mostly ignore each other



- State-action space grows linearly with number of agents
- Local communication, required only for coordination

"Interaction-Driven" Model

- Interaction Driven Markov Game[3]

$$\Gamma = (M_1, \ldots, M_n, {}^1M_I, \ldots, {}^mM_I)$$

Single agent MDPs ($M_1, \ldots, M_n$)

MMDPs (${}^1M_I, \ldots, {}^mM_I$)

  - Single-agent MDP for each agent;
  - MMDP for each "interaction area"

[3]M. Spaan, F. Melo. Local Interactions in Decentralized Multiagent Planning under Uncertainty. AAMAS, pp. 525-532, 2008.

Outline

- MMDPs vs. Dec-POMDPs
- Simplified Models
- Decoupling Interaction
- **Approximate Solutions**

IDMGs Revisited

- 2 components in IDMGs:
  - Single agent MDPs → Individual goals
  - Multiagent MDPs → Local interactions
- In "interaction" states, communication can be used to ensure coordination
- Communication is local

From IDMGs to Policies I

Case 1: Fully decoupled system

- IDMG Γ $= (M_1, ..., M_n)$
- Policy for agent k:

$$\pi_k^*(s_k) = \arg\max_{a_k} Q_k^*(s_k, a_k)$$

- $Q_k{}^*$ can be computed from M_k via DP

From IDMGs to Policies II

Case 2: Fully coupled system

- IDMG Γ $= (M_1, \ldots, M_n, M_I)$
- For each M_k, $r_k \equiv 0$
- Joint policy:

$$\pi^*(s) = \arg\max_a Q_I^*(s, a)$$

- Q_I^* can be computed from M_I via DP

From IDMGs to Policies III

General case:

- IDMG Γ = $(M_1, \ldots, M_n, {}^1M_I, \ldots, {}^mM_I)$
- For any (joint) policy:

$$Q_k^{\Gamma}(s,a) = Q_k^{\pi}(s_k, a_k) + \sum_i {}^iQ_I^{\pi}(s,a)$$

- Our approximation:

$$Q_k^{\Gamma}(s,a) \approx Q_k^{*}(s_i, a_i) + \sum_i {}^iQ_I^{*}(s,a)$$

Approximate Solution I

- Outside interaction states:

$$ {}^{i}Q_I^*(s,a) \approx 0 \qquad Q_k^{\Gamma}(s,a) \approx Q_k^*(s_i, a_i) $$

- Policy for agent k:

$$ \pi_k^*(s_k) = \arg\max_{a_k} Q_k^*(s_k, a_k) $$

Approximate Solution II

- In interaction states:
  - Agents have different interests:

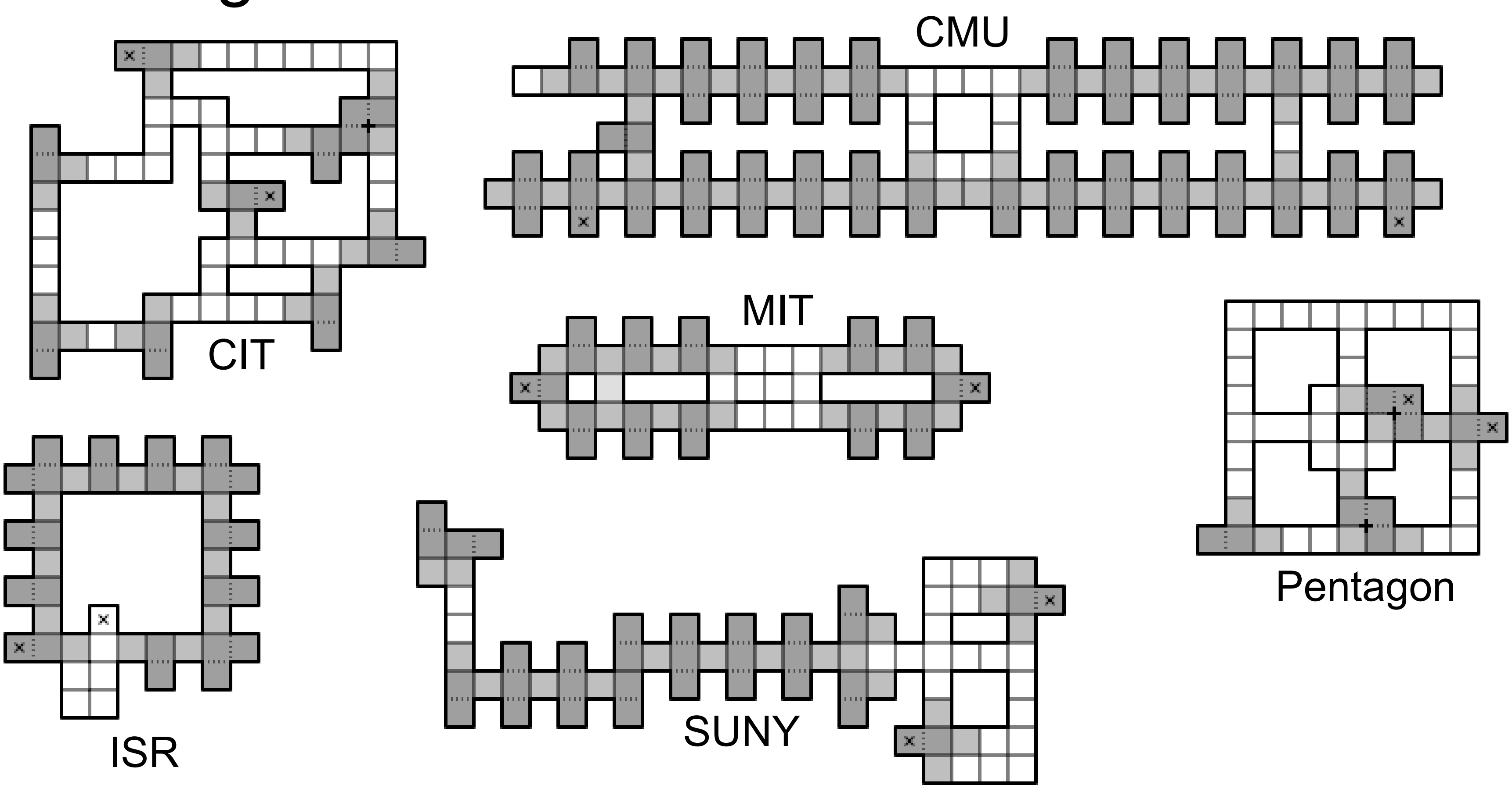
$$\arg\max_a Q_k^\Gamma(s,a) \neq \arg\max_a Q_l^\Gamma(s,a)$$

  - Game theoretic approach:

$$\pi_k^*(s) = \text{Nash}_a\left(Q_1^\Gamma(s,a),\ldots,Q_n^\Gamma(s,a)\right)$$

Example I

- Navigation scenarios:



Example I (cont.)

- Dimension of the environments:

Env.	Ind. States	Inter. States	Joint States
ISR	172	1,218	29,584
MIT	196	1,052	38,416
Pentagon	208	568	43,264
CIT	280	888	78,400
SUNY	296	1,668	87,616
CMU	532	4,574	283,024

Example I (conc.)

- Performance:

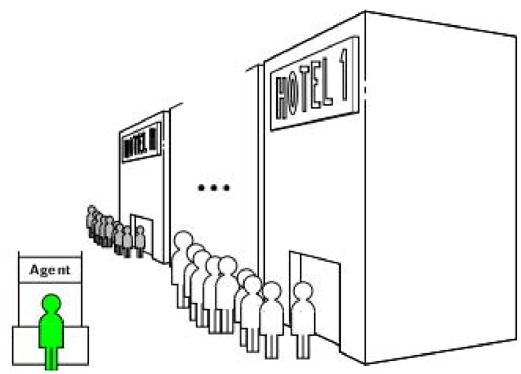
Env.	Indiv.*	IDMG*	Optimal**
ISR	-8.03	10.57	10.6
MIT	3.73	8.58	8.6
Pentagon	6.03	11.57	11.63
CIT	5.47	8.91	8.93
SUNY	-1.05	7.99	8.0
CMU	-3.16	5.81	5.84

Int. States
1,218
1,052
568
888
1,668
4,574

* The performance was obtained by averaging the total reward for the two agents.
** The optimal performance was obtained from an MMDP with joint reward the average of the individual rewards of the agents.

Example II

- The hotel problem:
  - Agent has available *N* rooms in each of *H* hotels;
  - At each time, client arrives for a random hotel;
  - The agent must assign clients to hotels;
  - Agent may also assign clients to luxury resort.



Example II (cont.)

- Dimension of the problem:

Env.	Ind. States	Inter. States	Joint States
$N = 1, H = 1$	4	3	16
$N = 2, H = 2$	27	180	729

Example I (conc.)

- Performance:

Env.	Indiv.*	IDMG*	Optimal**
N = 1, *H* = 1	58.92	**74.03**	87.63
N = 2, *H* = 2	97.84	**118.22**	133.85

Int. States
3
180

Env.	Indiv.*	IDMG*	Optimal**
N = 1, *H* = 1	58.92	**75.87**	87.63
N = 2, *H* = 2	97.84	**119.08**	133.85

Int. States
15
693

* The performance was obtained by averaging the total reward for the two agents.
** The optimal performance was obtained from an MMDP with joint reward the average of the individual rewards of the agents.

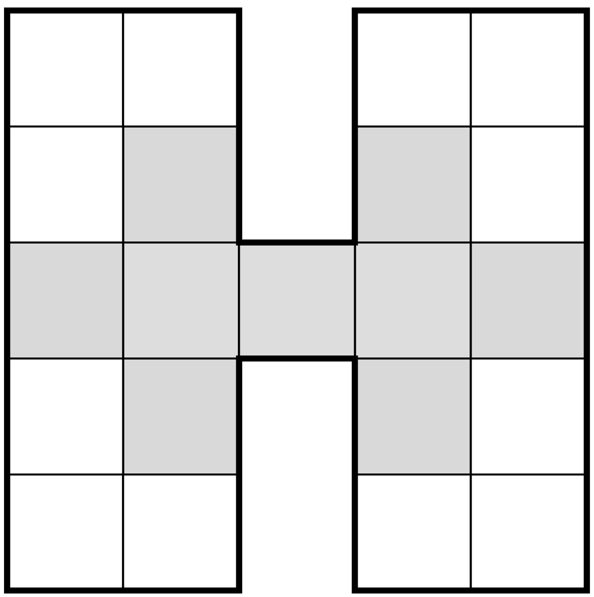
IDMGs Pros & Cons

- Advantages
  - Decouples interaction
  - Uses only local communication
  - Allows for agents with different interests
- Disadvantages
  - Depends on choice of “interaction” states
  - Requires computation of Nash (≈ NP-complete)[4]

[3]C. Papadimitriou. On the Complexity of the Parity Argument and Other Inefficient Proofs of Existence. J. Comp. Syst. Sci. 48(3), pp. 498-532, 1994.

Discussion

- Impact of "interaction states" on performance:



- Situation 1: No coordination is possible
- Situation 2: Coordination possible (greater delay)
- Situation 3: Coordination possible (smaller delay)

Conclusion

- Multi-robot/agent planning is complex
  - Communication
  - Observability
  - Scalability
- What if no “need” for “full” planning?
  - Limited interactions - MDPs
  - Interaction Driven Markov Games