Distributed Learning in Collaborative Control and Decision Making

John S. Baras

The Institute for Systems Research,
Departments of Electr. and Comp. Engin., BioEngin., Mech. Engin.,
Applied Mathematics, Statistics and Scientific Computation Program
University of Maryland College Park,
and
Guest Professor, School of Electr. Engin. and ACCESS Center, KTH,
and
TUM-IAS Hans Fischer Senior Fellow

Inaugural Lecture as HF SF

October 21, 2014
Institute for Advanced Study, Technical University of Munich
Munich, Germany

Acknowledgments

- **Joint work with:** Pedram Hovareshti, Anup Menon
- **Sponsors:** NSF, AFOSR, ARL, DARPA, NIST

Outline

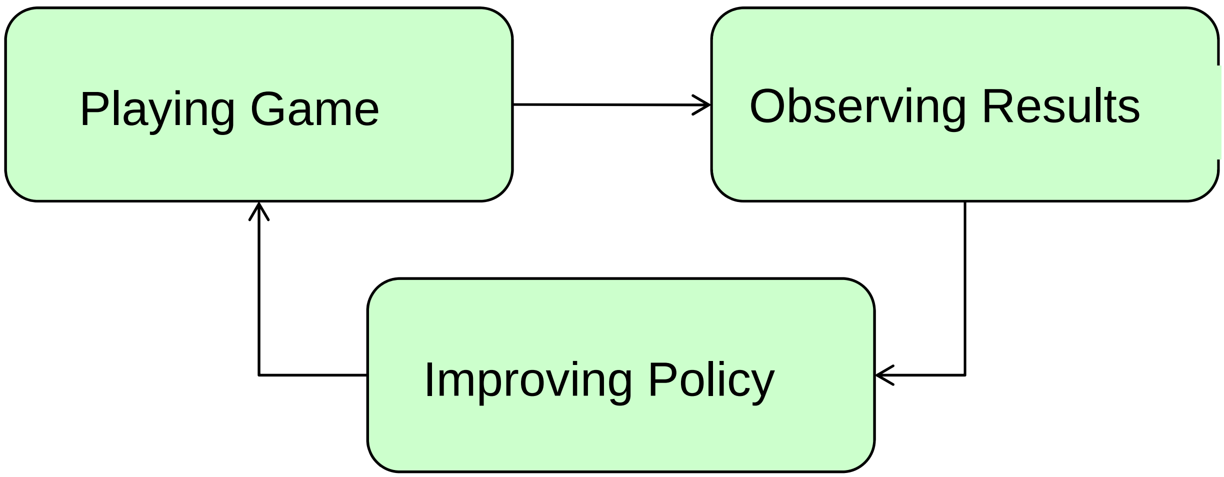
- Autonomous decision making and learning
- Coordination games
- Types and behaviors –learning
- Analysis – convergence
- Simulations
- Distributed learning from interactions: examples
- Problem formulation (discrete action space)
- How learning in repeated games can help
- Modeling framework and a simple algorithm
- Problem formulation (continuous action space)
- Modeling framework and extremum seeking control
- Wind farm management simulations
- Conclusions and future work

Autonomous Decision Making

- When making a decision, an agent is influenced by its knowledge about the other agents' behavior
- **Problem**: ***Modeling decision making on whether to cooperate in a group effort as a result of two person games on a network***
- Adaptation to neighbors' strategies as a coordination mechanism
- The system is analyzed under classes of linear and bounded linear behavior functions; A generalized consensus problem determines strategy coordination
- The ***emerging collaboration*** graph is a function of agents' behavioral tendencies as well as the connectivity graph

Motivation: Learning in Games

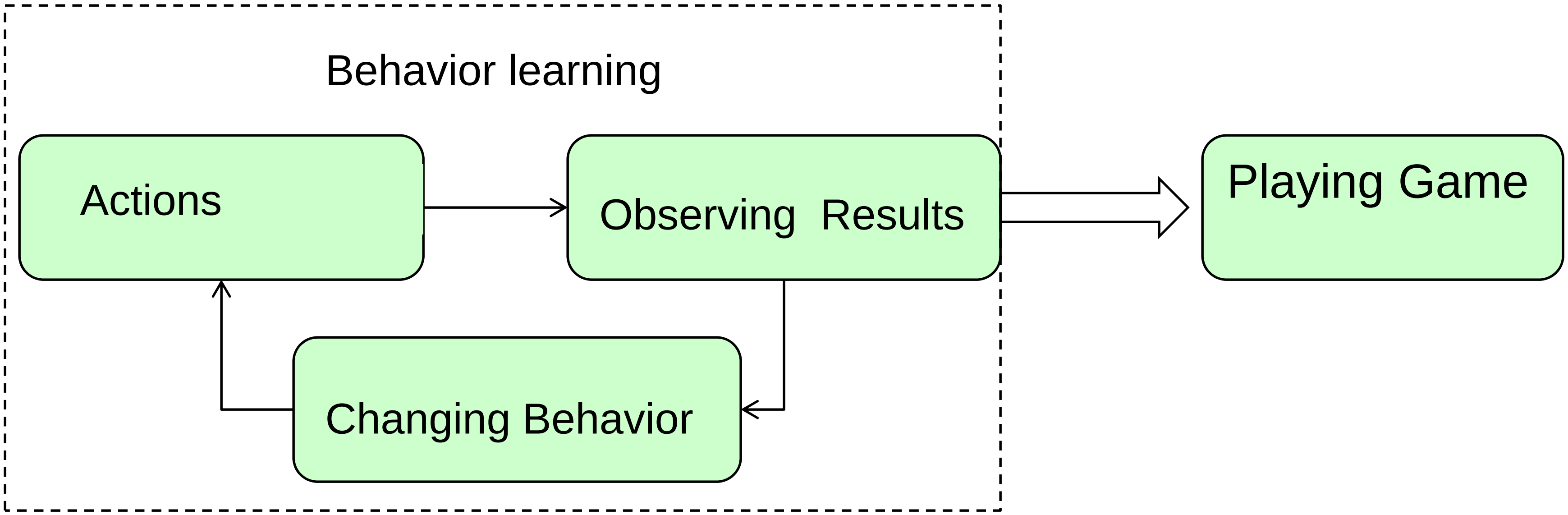
- To explain why equilibrium arises as the long run outcome with non-fully-rational players



- Acceptable results in long run repetitive situations
- What about one shot and short term games that rely heavily on players prior beliefs about each other?
- We address the problem of learning to coordinate for a one-time situation

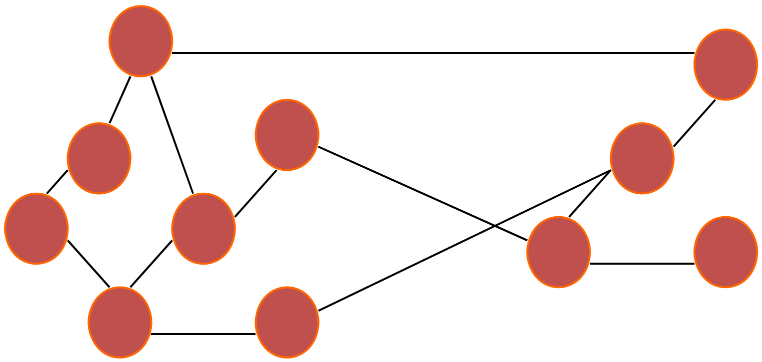
Learning to Coordinate

- Agents to decide on whether to participate in a collaborative effort based on their understanding of others' tendencies and what they believe that others' understand about their neighbors tendencies and …



- Example: whether or not to take part in a riot
- ***Emergence of a collaboration graph from communication***

System Model



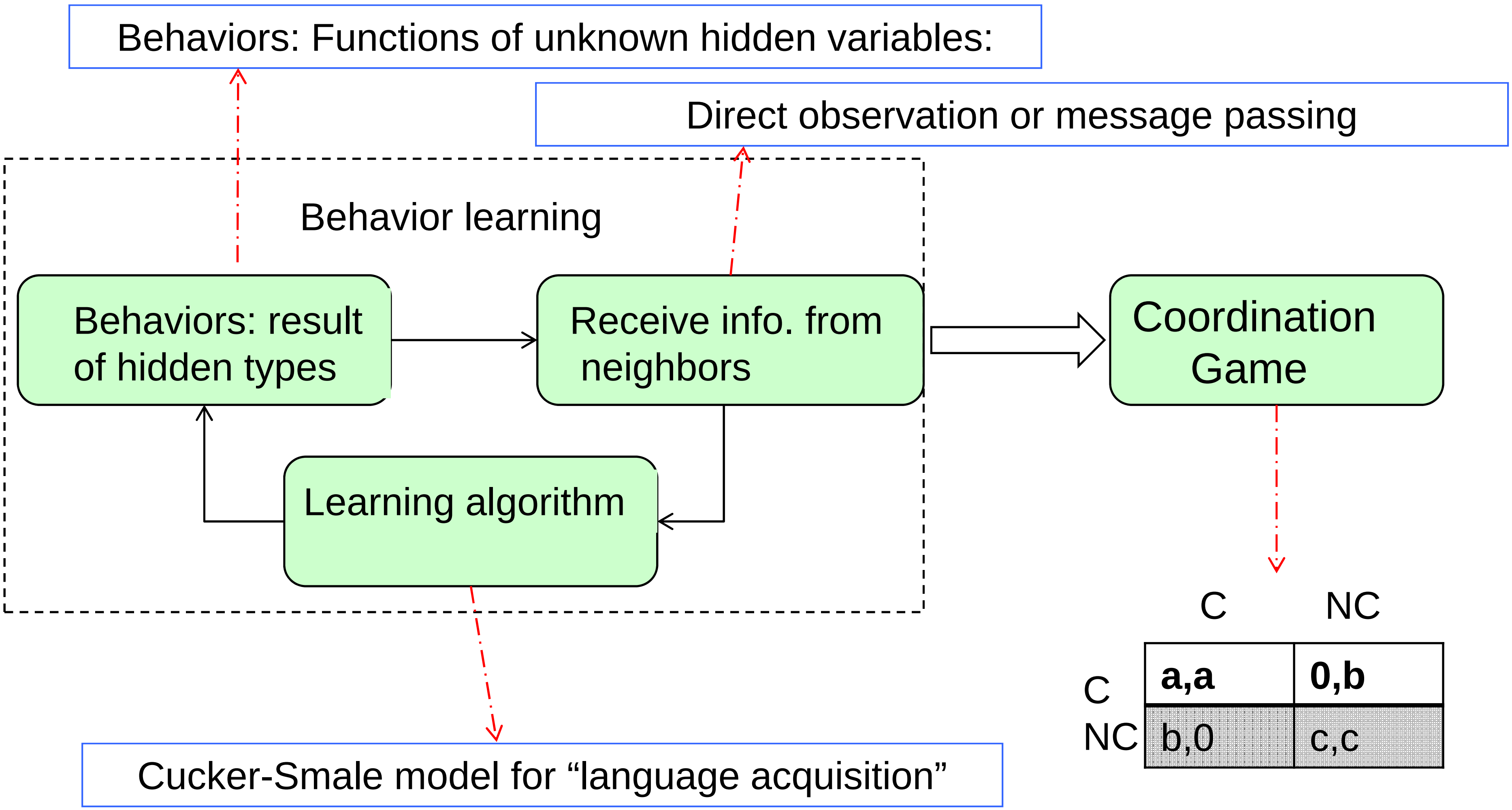
$$G = (V, E)$$

$$V = \{1, 2, \ldots, n\}$$

$$E \subseteq V \times V$$

- Each agent has to make a decision on whether to cooperate (*C*) or not (*NC*) in a group effort
- Based on its decision it will incur a payoff which is the sum of payoffs resulting from playing 2-person coordination games with all neighbors
- Agents strategy based on their type
- Agents learn and adapt to neighbors' strategies modeled in Cucker-Smale framework

System Model Overview



	C	NC
C	**a,a**	**0,b**
NC	b,0	c,c

$$a > b > c > 0.$$

The Coordination Game

	C	NC
C	**a,a**	**0,b**
NC	b,0	c,c

$$a > b > c > 0.$$

- ***Cooperation is the Pareto-optimal*** equilibrium strategy, whereas ***Not Cooperation is the risk sensitive one***
- Agent payoff is sum of its 2-person games payoffs with its neighbors

$$u_i(s_i, s_{-i}) = \begin{cases} a \sum_{j \in N_i} 1_{\{s_j = C\}} & \text{if } s_i = C \\ b \sum_{j \in N_i} 1_{\{s_j = C\}} + c \sum_{j \in N_i} 1_{\{s_j = NC\}} & \text{if } s_i = NC \end{cases}$$

Types and Behaviors

- Each agent has a behavior system that decides on its *level of optimism* (playing C)
- This system evolves in time: Cucker-Smale framework for language evolution
- *Behavior* (or type): A function

$$f : X = [0,1] \rightarrow Y = [0,1]$$

- Given a uniformly distributed RV, *x*, f_i determines whether agent *i* expects an event that is supposed to occur with probability *x*, to actually happen

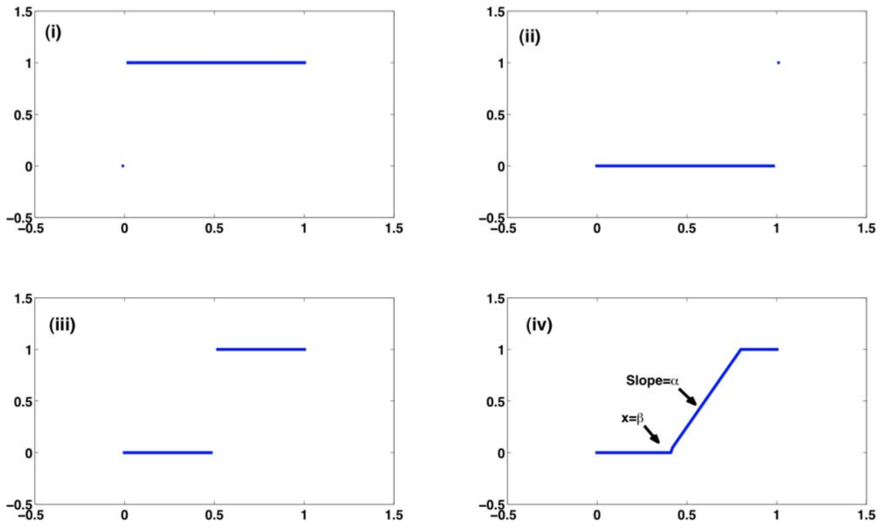
Types and Behaviors

Optimist

Pessimist

Ambivalent

Regular



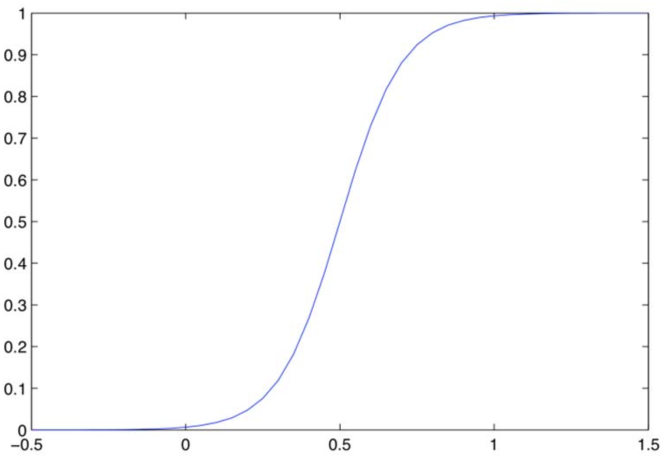
Types are modeled by a set of functions, e.g.

$\mathbf{F}$: The set of sigmoids with following property:

$$f(x;\theta_1,\theta_2) = \frac{1}{2}\left[1+\tanh(\theta_1(x-\theta_2)\right],$$

$$\theta_{1\min} \le \theta_1 \le \theta_{1\max},$$

$$\theta_{2\min} \le \theta_2 \le \theta_{2\max}.$$



Learning Infrastructure

- Agents learn and adapt to neighbors' types
- Given a communication infrastructure, the neighbors' ***influence*** and ***interaction*** is modeled using a stochastic matrix

$$W = [w_{ij}],$$

$$\sum_j w_{ij} = 1,$$

$$(i,j) \notin E \Rightarrow w_{ij} = 0.$$

Relative Influence of node j on node i.

- ***W*** is a measure of influence and trust

Learning Algorithm

- A version of Cucker-Smale algorithm for "language acquisition"
  - At each time each agent *i* receive neighbors data

$$\{x_j(t), y_j(t) = f(x_j(t), \theta_1, \theta_2)\}_{j \in N(i)}$$

Set of parametrized functions F

Distributed uniformly on $X = [0,1]$

  - Agents update their type function as:

$$f(x_i(t+1)) = \arg\min_{f \in \mathbf{F}} \sum_{j \in N(i)} w_{ij} (f(x_j(t) - y_j(t))^2,$$

$$i = 1, 2, \ldots, n$$

Analysis for Linear Behavior Functions

- Class of bounded linear functions

$$F_l^* = \{f \mid f(x) = \theta x + \lambda;\ \theta \in [\theta_{\min}, \theta_{\max}];\ \lambda \in [\lambda_{\min}, \lambda_{\max}]\}$$

- Class of linear functions

$$F_l = \{f \mid f(x) = \theta x + \lambda;\ \ \theta, \lambda \in \mathrm{R}\}$$

- ***Theorem***: If all agents use bounded linear behavior functions, the learning algorithm converges with probability 1 to a consensus on behavior functions, provided that the matrix *W* is irreducible.

Relaxing Boundedness Assumption

- Using linear assumption system evolves as

$$\Theta(t+1) = \begin{bmatrix} P_1(t) & M_1(t) \\ M_2(t) & P_2(t) \end{bmatrix} \Theta(t)$$

$$P_1 1_n = 1_n, \quad M_1 1_n = 0,$$
$$M_2 1_n = 0, \quad P_2 1_n = 1_n,$$

in which

$$\Theta = \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_n & \lambda_1 & \lambda_2 & \dots & \lambda_n \end{bmatrix}^T$$

- Reaching consensus in this setting requires ***consensus on both variables***

Convergence Theorems

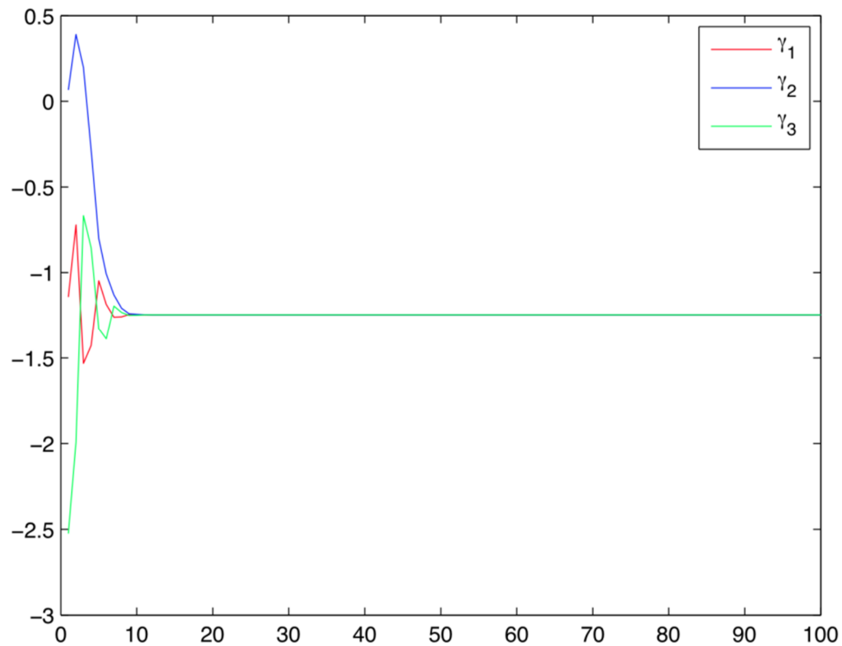
- For the one time learning case, the agents will reach a consensus on θ and γ with probability 1, i.e. they will coordinate on the same behavior function $f(x) = \theta^* x + \gamma^*$, θ^* and γ^* are the fixed points of

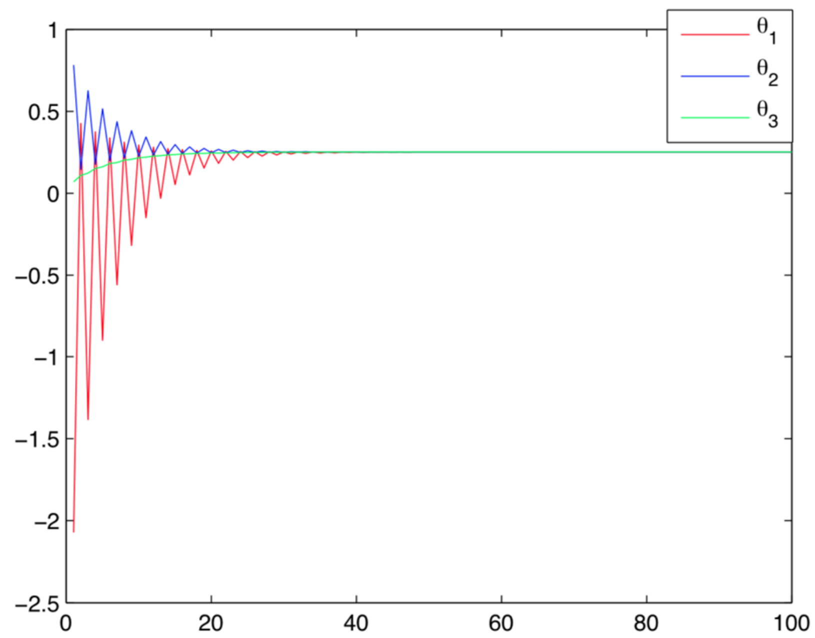
$$\Theta(t+1) = \begin{bmatrix} P_1(t) & M_1(t) \\ M_2(t) & P_2(t) \end{bmatrix} \Theta(t)$$

- In the general case, the agents will reach a consensus on the behavior function with probability 1

Simulations

- Game model with a=5, b=4, c=2
- Runs for 3 agent complete networks
- Fast convergence of θ and γ, the strategy parameters





Further Observations

- In cases with ***majority of agents optimistic***, ***optimist*** behavior emerges
- In cases with minority of agents optimistic, optimist behavior can also emerge

Example – Windfarms[a, b]

Horns Rev 1 wake effects. Courtesy Christian Steiness

- No good models for aerodynamic interactions between turbines.
- Objective – maximize total power production.

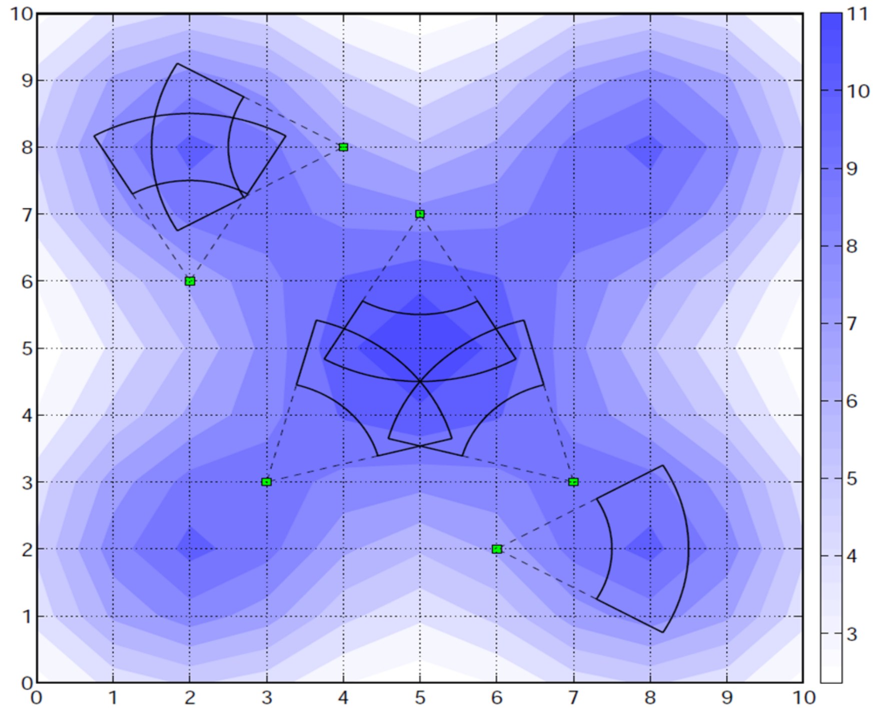
Assign individual utility

$u_i(t) =$ power produced by turbine *i* at time *t*

such that maximizing $\sum_i u_i(t)$ leads to desirable behavior.

[a]. Gebraad, van Dam, and van Wingerden, "A model-free distributed approach for wind plant control," ACC, 2013.
[b]. Marden, Ruben, and Pao, "A model-free approach to wind farm control using game theoretic methods," IEEE Trans. Control Systems Tech, 2013.

Example – Source Seeking, Coverage[c]



Darker the shade of blue, more the interest in the site. Sectors represent sensor position.[c]

Design individual utility

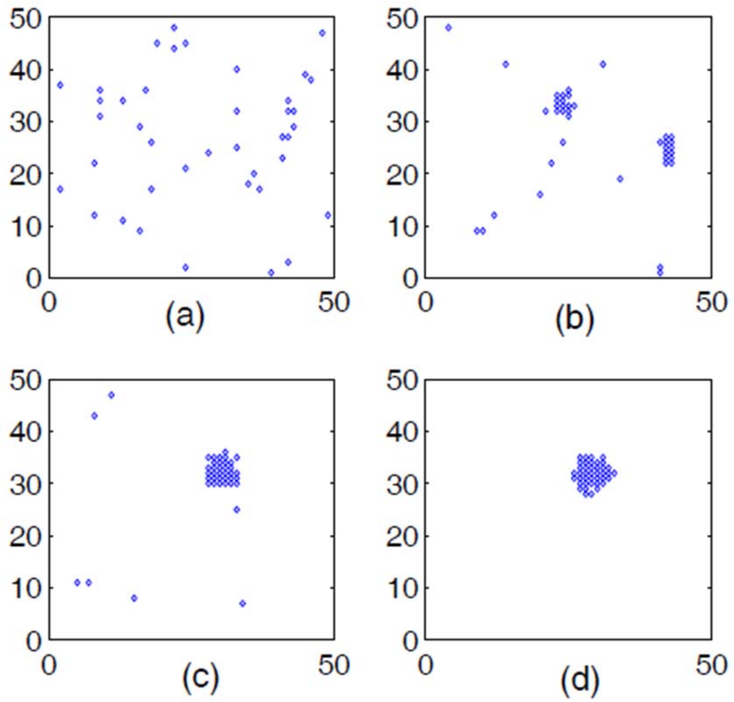
$$u_i(s, c) = \sum_{s' \in NB(s,c)} \frac{q(s')}{n(s')} - f_i(c),$$

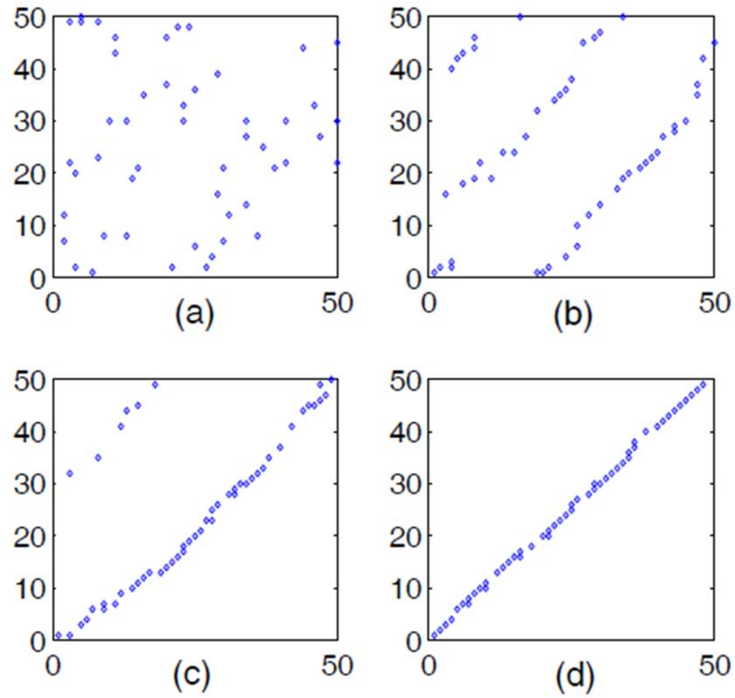
such that maximizing $\sum_i u_i(t)$ leads to desirable behavior.

(here q(s)= interest in observing s, n(s) = number of agents observing s, NB(s,c) = subset of S observable from s when camera viewing angle= c, and $f_i(c)$ = processing cost when the camera viewing angle is c.)

[c]. Zhu and Martinez, "Distributed coverage games for energy-aware mobile sensor networks," *SIAM J. on Control and Optimization*, 2013.

Example – Formation Control[d, e]





Simulation results demonstrating rendezvous and gathering along a line[a]

For rendezvous, design individual utility

$$u_i(s_i) = \frac{1}{|\{s_j \in S : \|s_i - s_j\| < r\}|} - \alpha \, dist_{\leq r}(s_i, obstacle),$$

such that minimizing $\sum_i u_i(t)$ leads to desirable behavior.

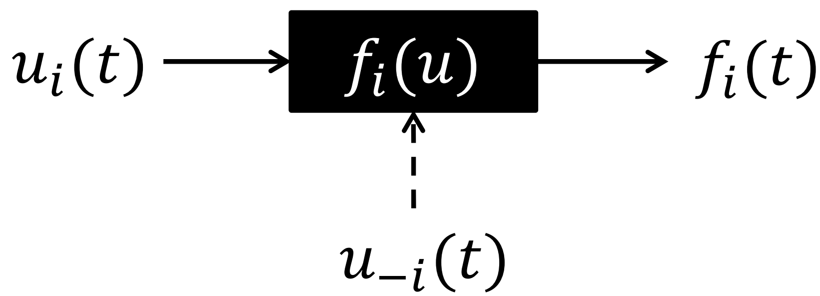
[d] Xi, Tan and Baras, “Decentralized coordination of autonomous swarms using parallel Gibbs sampling," *Automatica*, 2010.
[e]. Baras et al., “Decentralized Control of Autonomous Vehicles," *Proc. of IEEE CDC*, 2003.

Problem Formulation

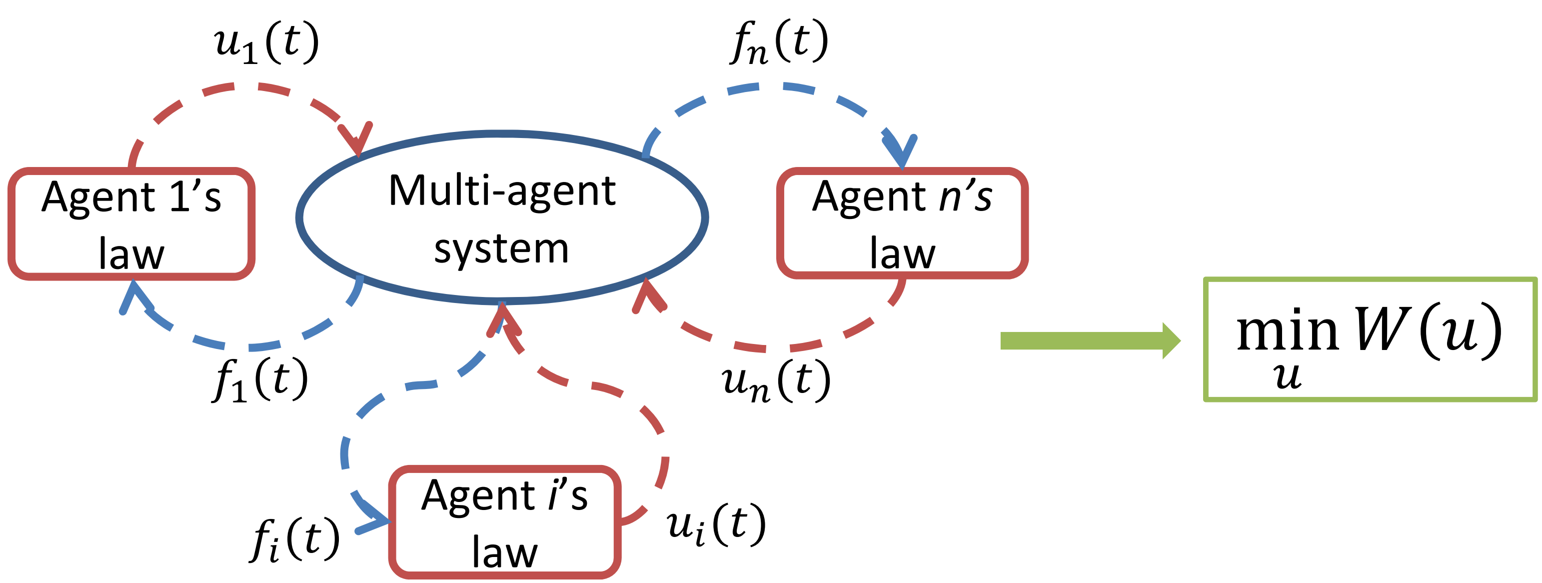
Engineered Multi-agent System

- n agents
- Utility fn. $\{f_1(u), \ldots, f_n(u)\}$
- Welfare fn. $W(u) = \sum_i f_i(u)$

Model Free Set-up



Collaborative Objective



Formulation (discrete action space)

- N agents, agent i picks actions from a finite set A_i.

- Agent i receives/measures private utility
$$u_i : A \rightarrow R^+$$
where $A = \prod_{i=1}^{N} A_i$.

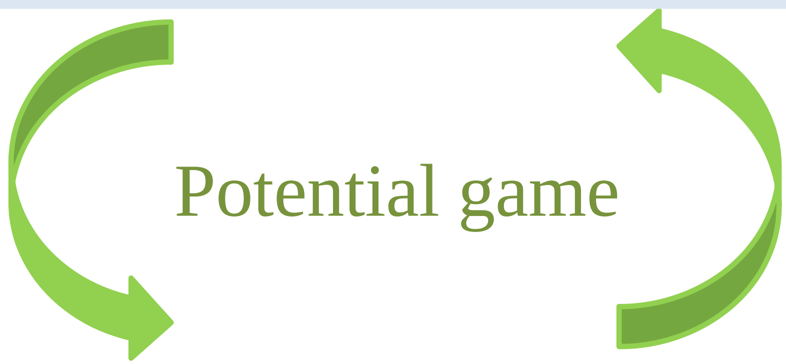
- Minimize W(a) = $\sum_{i=1}^{N} u_i(a)$ over A → seek the efficient actions
$$A^* = \{\underset{a \in A}{\operatorname{argmin}} W(a)\}.$$

- Agent knows past actions and payoffs – $\{(a_{t-1})_i, (u_{t-1}^{mes})_i, \ldots, (a_0)_i, (u_0^{mes})_i\}$.

Approach using Learning in Games

1. **Utility assignment**
such that solution concepts like Nash Eq. (NE) in resulting 'game' correspond to desirable system-wide outcomes.

In potential games, "efficient outcomes" correspond to NE.



Most learning rules converge to NE for games with special structure.

2. **Prescribe Learning Rule**
for agents to learn equilibria.
Ex. log-linear learning, fictitious play, adaptive play, regret-matching etc.

Example Application – Consensus Problems[f]

A potential game is one where there exists a function φ such that
$u_i(a_i, a_{-i}) - u_i(a'_i, a_{-i}) = \varphi(a_i, a_{-i}) - \varphi(a'_i, a_{-i}) \forall\, i.$

In a potential game, maximizer of φ correspond to NE .

Consider N (non-strategic) agents each with a discretized set of actions; A_i for *i.*

Assign utility $u_i(a) = -\sum_{j \in N_i} ||a_i - a_j||$ → computable from local measurements.

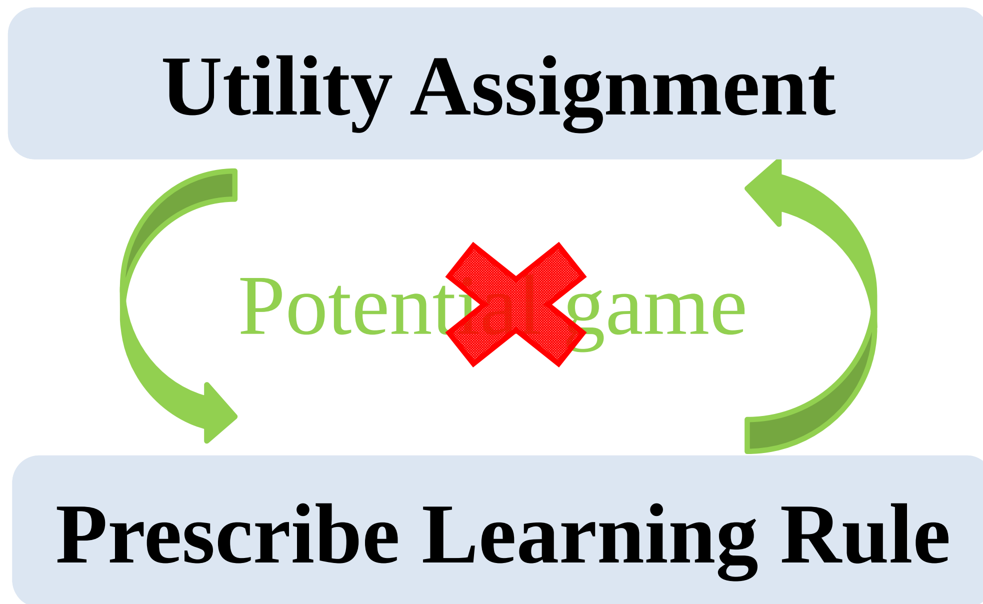
The resulting 'game' is a potential game with potential function

$$\varphi(a) = -\sum_{i} \sum_{j \in N_i} \frac{1}{2} ||a_i - a_j||.$$

Program agents to follow a `learning rule' → consensus.

[f]. Marden, Arslan and Shamma, "Cooperative Control and Potential Games," *IEEE Tran. on System, Man and Cybernetics*, 2009.

Shortcomings



Not always possible to assign utilities with special structure!

➡ NE may be inefficient.

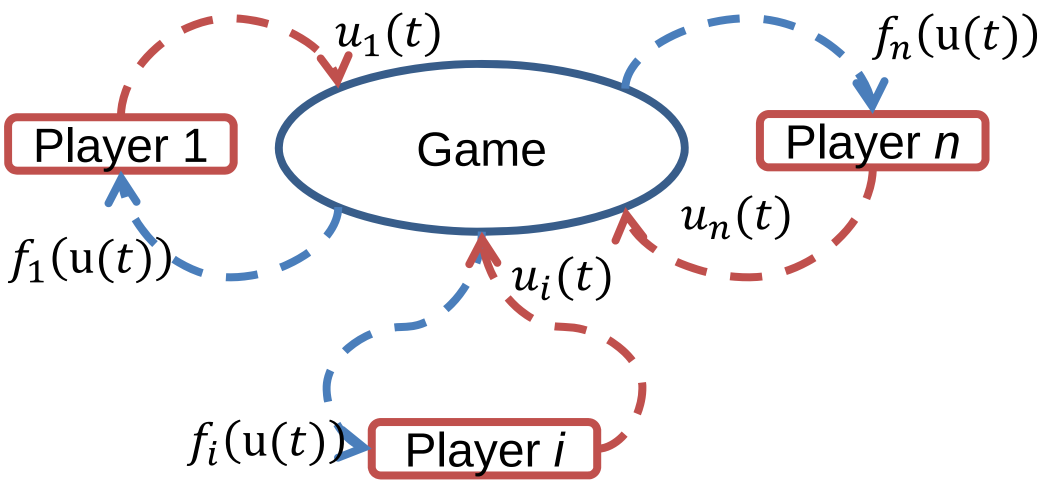
➡ Known learning rules needn’t converge.

Desired Features

- Payoff-based implementation.
- Solution concept – welfare optimality.
- Converges regardless of utility structure.

Learning Rule	Utility Assumption	Implementation
Fictitious Play	Potential Games	Excessive
Reinforcement L.	Common Interest	Payoff based
Adaptive play	Weakly Acyclic	Excessive
Log-linear L.	Potential Games	Excessive
Trial and Error L.	NE	Payoff based
Pradelski, Young	Eff. NE, ‘interdependence’	Payoff based
Marden, Young, Pao	Welfare max., ‘interdependence’	Payoff based

Learning in Games



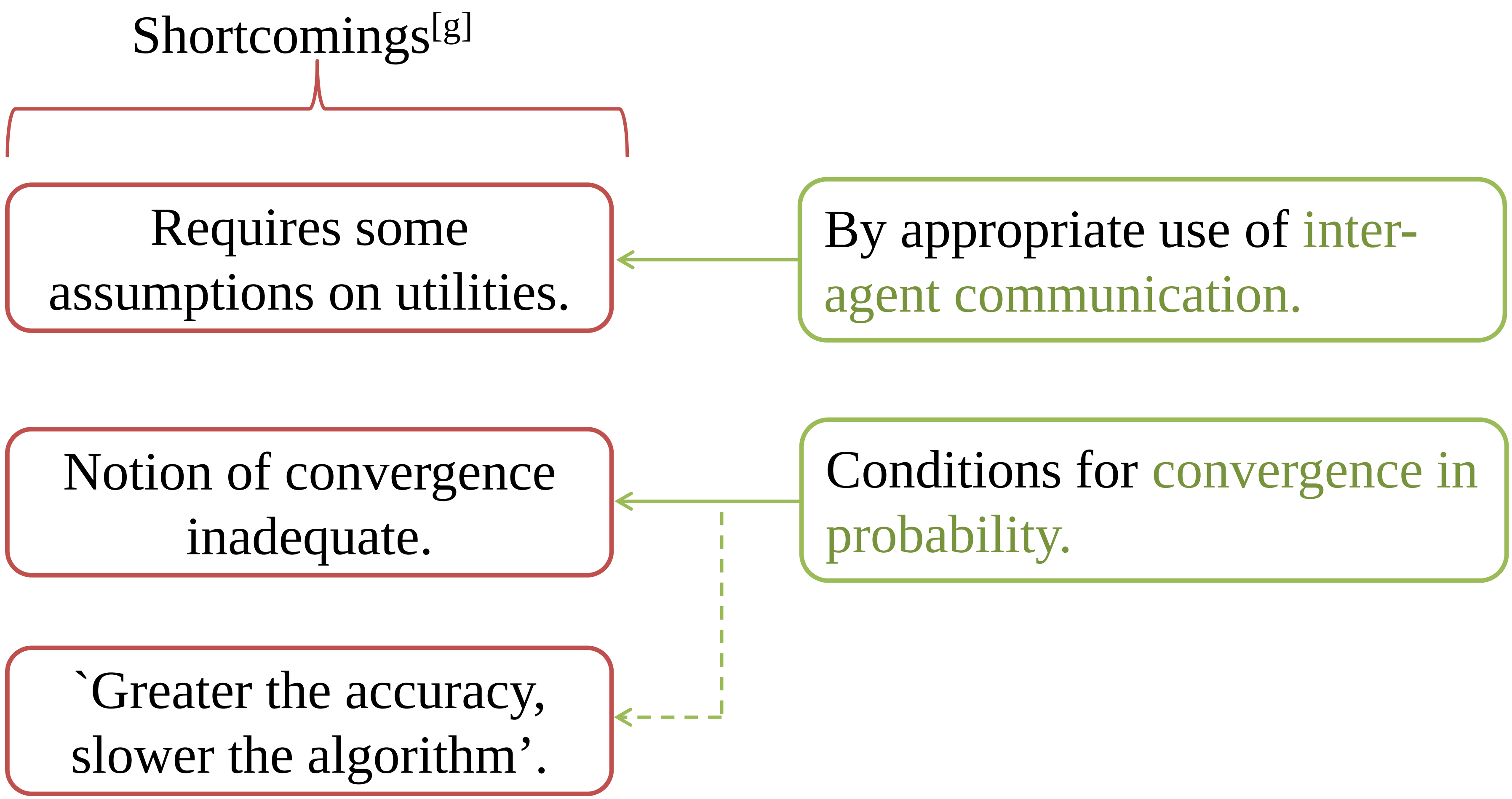
Learning Rule	Utility Assumption	Implementation
Fictitious Play	Potential Games	Excessive
Reinforcement L.	Common Interest	Payoff based
Adaptive play	Weakly Acyclic	Excessive
Log-linear L.	Potential Games	Excessive
Trial and Error L.	NE	Payoff based
Pradelski, Young	Eff. NE, 'interdependence'	Payoff based
Marden, Young, Pao	Welfare max., 'interdependence'	Payoff based

Simple "payoff-based" adaptation rules lead to interesting emergent behavior.

The Meta Theorem: When players adopt [learning rule] and if the game satisfies [property], then player actions converge to [equilibrium].

Beyond Nash equilibration → Converge to Welfare optimal actions without any assumptions on utilities (or "game").

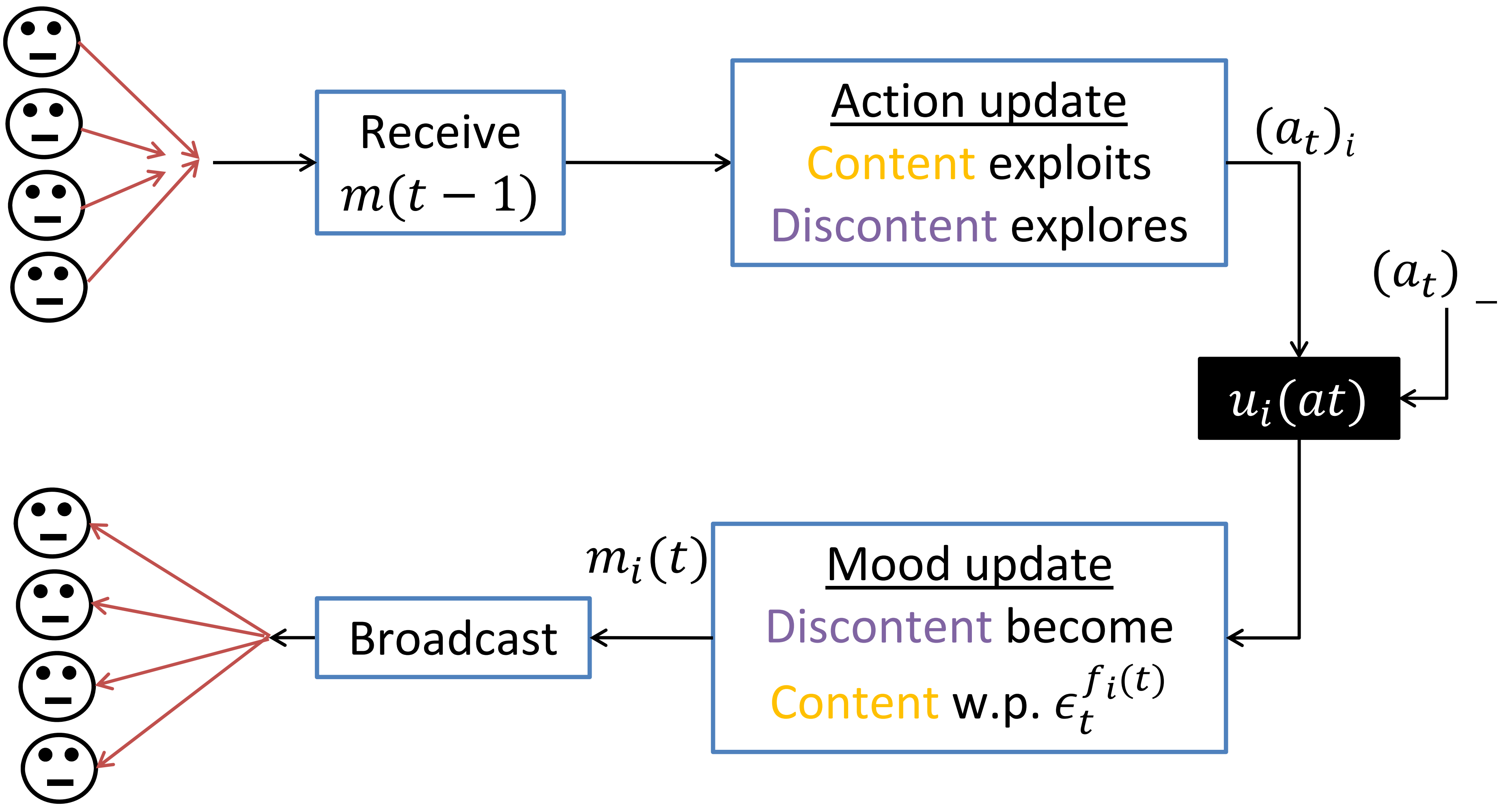
Our Contribution



[g]. Marden, Young and Pao. "Achieving Pareto optimality through distributed learning," *Proc. of IEEE CDC*, 2012.

Proposed Algorithm

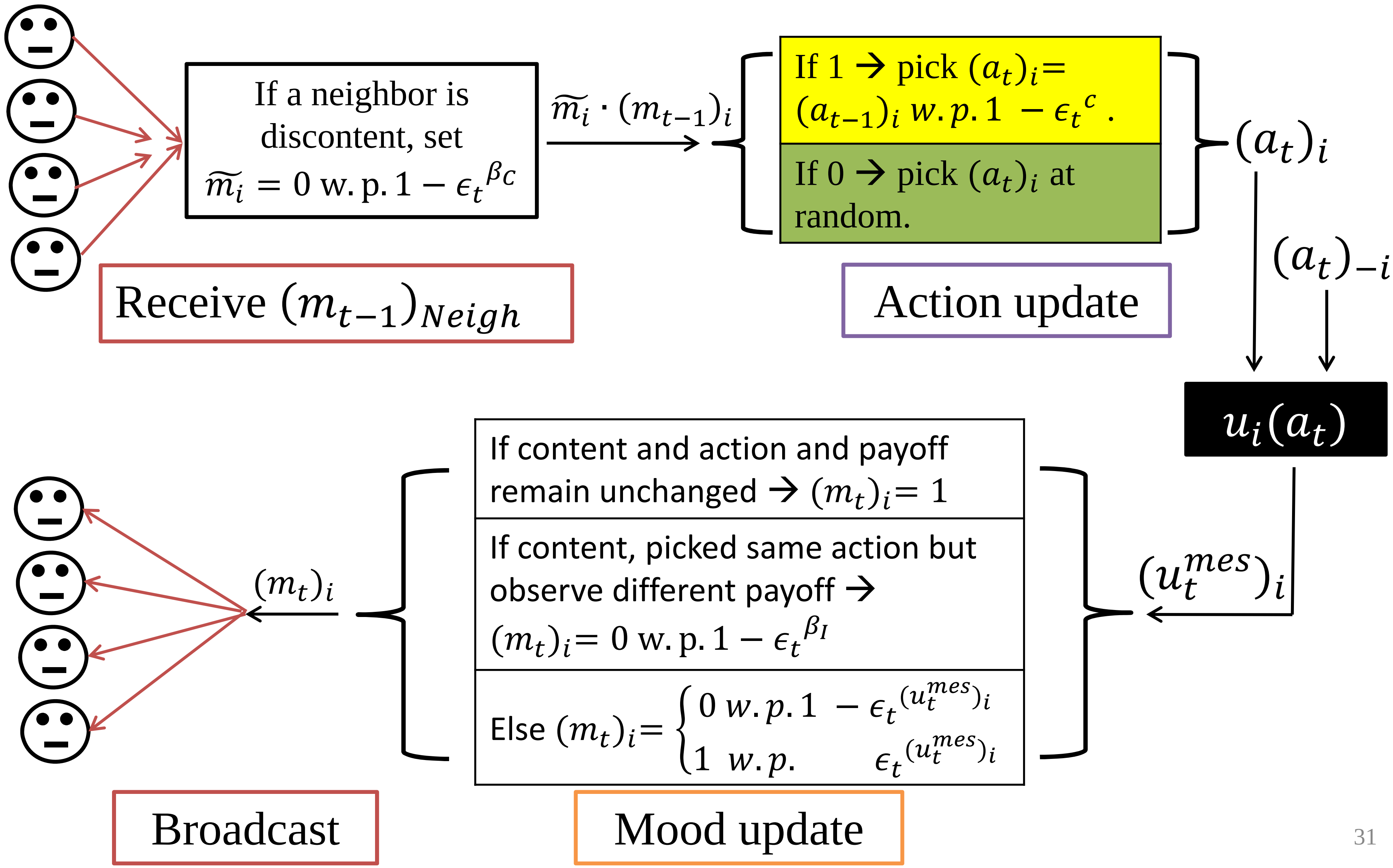
State $x_i = (u_i, m_i)$; $m_i = 1 \leftrightarrow$ content and $m_i = 0 \leftrightarrow$ discontent.



[g]. Marden, Young, Pao, "Achieving Pareto optimality through distributed learning," *IEEE CDC, 2012*.
[h]. Menon, Baras, "A distributed learning algorithm with bit-valued communications for multi-agent welfare optimization", IEEE CDC, 2013.

Proposed Algorithm (detail)

State $x_i = (a_i, m_i)$; $m_i = 1 \leftrightarrow$ content and $m_i = 0 \leftrightarrow$ discontent.

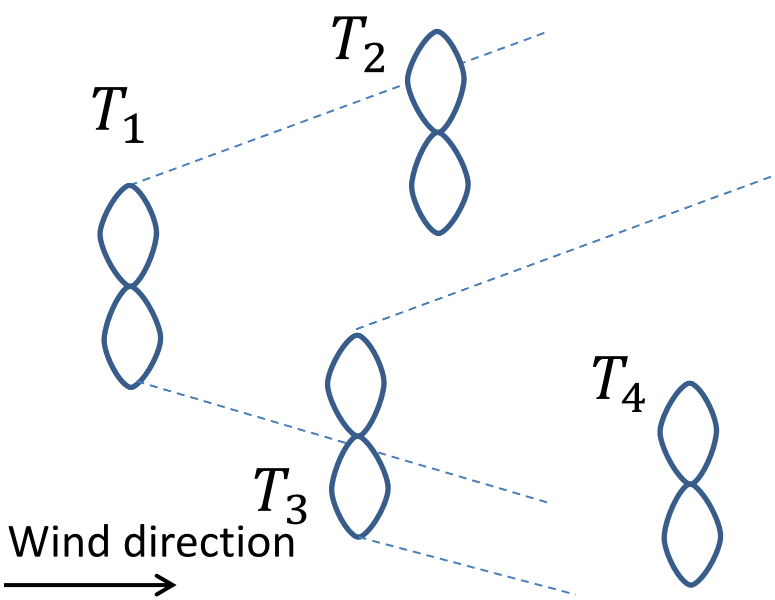


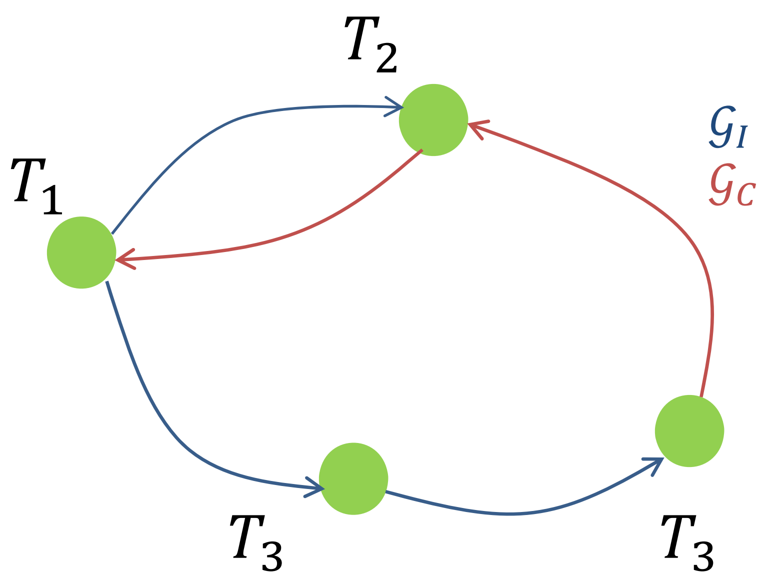
A Coarse Modeling Framework

Like agents, system designer doesn't know functional form of payoffs.

Interaction graph $\mathcal{G}_I$ models implicit communications:
Link (i, j) implies i's actions affect j's payoff.

Communication graph $\mathcal{G}_C$ models explicit communications:
Link (i, j) implies msg. sent by i is received by j.



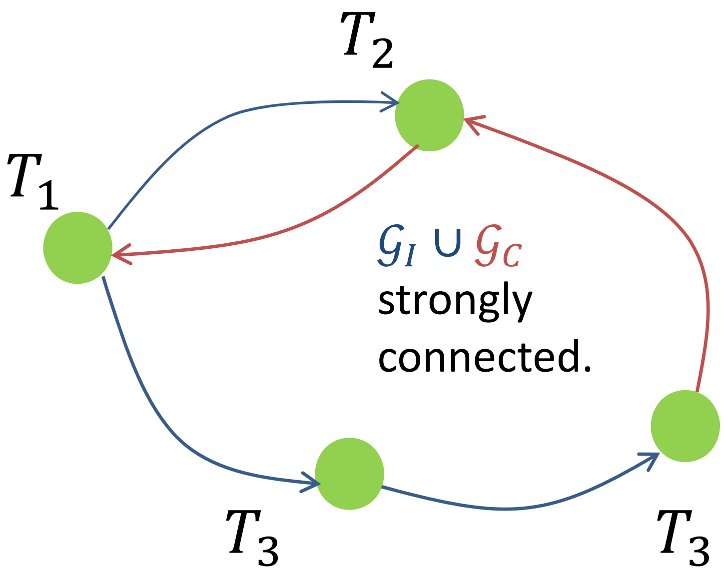


Convergence Guarantee

Theorem. *Assume* $c > W^*, \beta_I > 0, \beta_C > 0,$
1. for each $a \in A, \mathcal{G}_c(a) \cup \mathcal{G}_I(a)$ *is strongly connected and*
2. $\sum_{t=1}^{\infty} \varepsilon_t^c = \infty$
Then,

$$\lim_{t \to \infty} P(a_t \in A^*) = 1.$$

- The algorithm is model free – if nothing is known about $\mathcal{G}_I$, design $\mathcal{G}_C$ strongly connected.
- Communication is only bit-valued: simple implementation.



Proof Overview

Fix $\varepsilon_t \equiv \varepsilon > 0$. Algorithm is an irreducible, aperiodic Markov chain $P(\varepsilon)$; $\mu(\epsilon) = \mu(\epsilon)P(\epsilon)$.

$\lim_{\epsilon \to 0} \mu(\epsilon) = \mu(0)$ s.t. $\mu(0) = \mu(0)P(0)$.

→

If, $\mathcal{G}_c \cup \mathcal{G}_I$ is strongly connected, $\mu(0)$ has support over states with a $\in A^*, m_i = 1\ \forall\ i$.

↓ let ε vary as ε_t.

$\epsilon_t \to 0$ as $t \to \infty$. Nonhomogeneous Markov chain $\mathbf{P}(\mathrm{t}) = P(\varepsilon_t)$.

← Ensuring ergodicity.

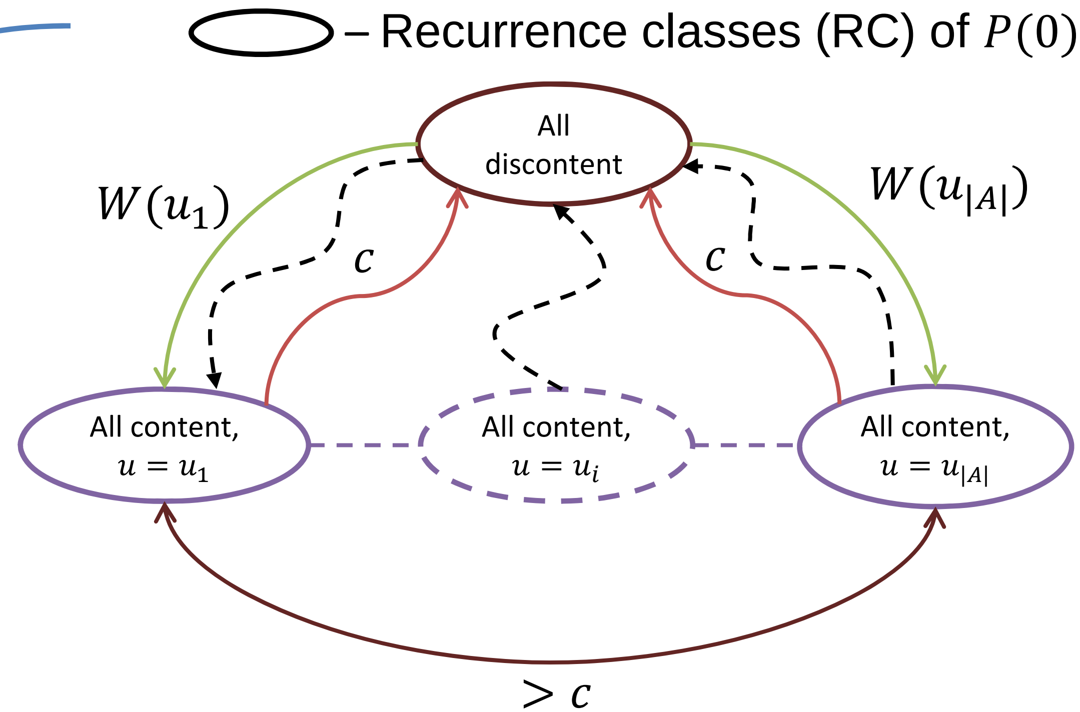
Rate condition $\sum_{t=1}^{\infty} \varepsilon_t^c = \infty$ ensures ergodicity of $\mathbf{P}(\mathrm{t})$ with $\mu(0)$ as limiting distribution.

Proof Overview

Step 1: Freeze $\epsilon_t \equiv \epsilon \rightarrow$ Irreducible, aperiodic Markov chain $P(\epsilon)$

Step 2: Stationary distribution of $P(\epsilon)$ for small $\epsilon > 0$?

Step 3: "Annealing" ϵ_t to 0 preserving ergodicity.



- RC with least resistive trees rooted at them are stochastically stable[a].
- Recall $c > W^* \rightarrow$ for the algorithm, the stochastically stable RC is where all agents are content and $u \in A^*$.

[i]. Young, "Evolution of Conventions", Econometrica, 1993.
[j]. Menon, Baras, "Convergence Guarantees for an Algorithm Achieving Pareto optimality", Proc. of ACC 2013.
[h]. Menon, Baras, "A distributed learning algorithm with bit-valued communications for multi-agent welfare optimization", CDC, 2013.

Ergodicity for time-varying Perturbed Markov Chains

Main Result: Ergodicity of nonhomogeneous Perturbed Chains [a]

Let the recurrence classes of the unperturbed chain P(0) be aperiodic and the parameter ε be scheduled according to the monotone decreasing sequence $\{\varepsilon(t)\}$, *with* $\varepsilon(t) \rightarrow \infty$ *as* $t \rightarrow \infty$. *Then, a sufficient condition for weak ergodicity of the resulting chain is*

$$\sum_t \varepsilon(t)^\gamma = \infty .$$

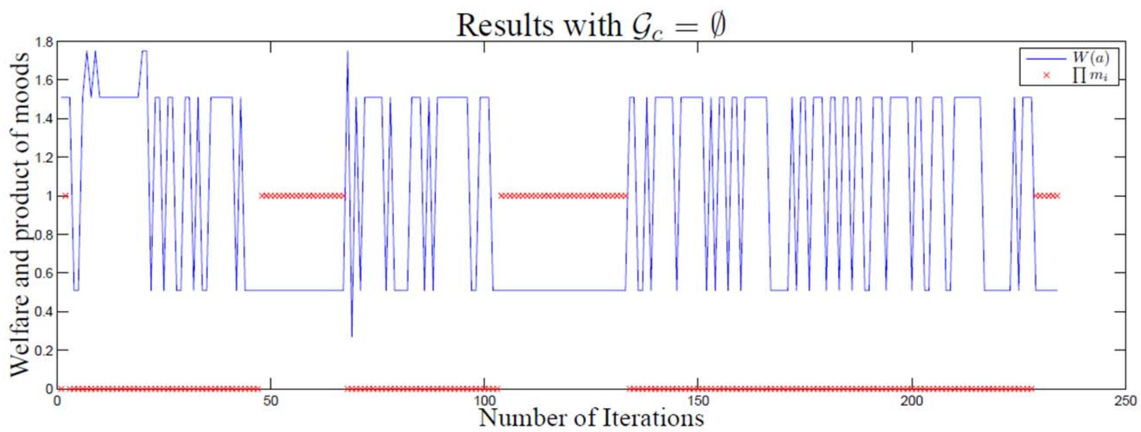
Furthermore, under mild assumptions on the structure of the transition probabilities, if the chain is weakly ergodic then it is strongly ergodic with the same limiting distribution μ(0) as described earlier.

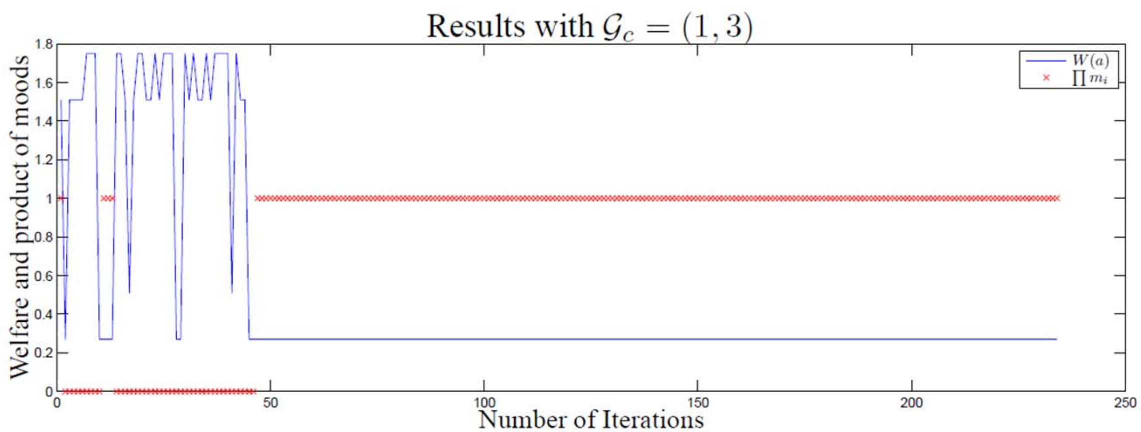
[j]. Menon and Baras, "Convergence Guarantees for an algorithm achieving Pareto optimality," *Proc. of ACC 2013.*

Simulations – Verifying Results

Agent 3 →	l	l	h	h
Agent 2 →	l	h	l	h
Agent 1				
l	$(\frac{1}{10}, \frac{1}{10}, \frac{1}{4})$	$(\frac{1}{2}, 1, \frac{1}{4})$	$(\frac{3}{4}, \frac{3}{4}, \frac{1}{10})$	$(1, \frac{1}{2}, \frac{1}{10})$
h	$(1, \frac{1}{2}, \frac{1}{4})$	$(\frac{3}{4}, \frac{3}{4}, \frac{1}{4})$	$(\frac{1}{2}, 1, \frac{1}{10})$	$(\frac{1}{4}, \frac{1}{4}, \frac{1}{10})$

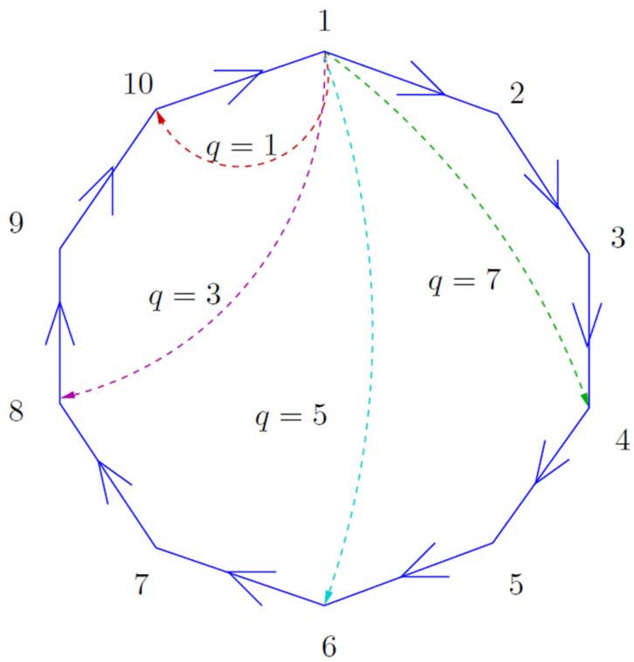
Payoff structure of a three-agent system

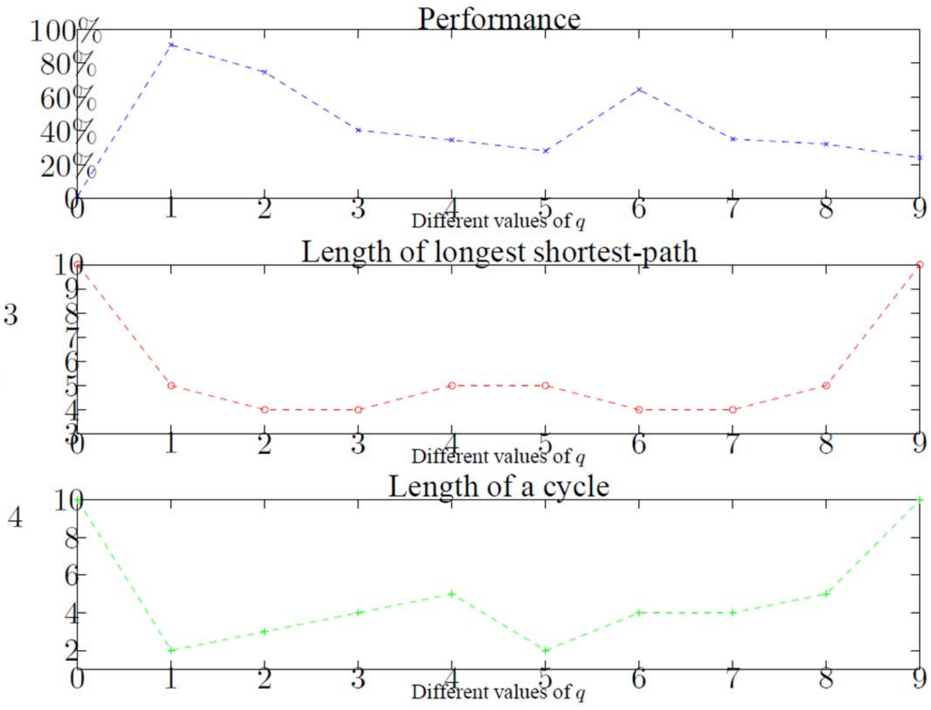




Simulations – Dependence on $\mathcal{G}_c$

- N agents, $A_i = \{0.1,1\}\ \forall\ i$.
- $u_i(a) = a_{i-1}$,
- $\mathcal{G}_c^q$ has edges $(i, i-q)$.





Simulation results for $N = 10$.

Simulations – Dependence on $\mathcal{G}_I$

- N agents, $A_i = \{0.1,1\}\ \forall\ i$.
- $u_i(a) = \frac{1}{1+2q}\sum_{j=i-q}^{i+q} a_j$ (index ops. mod N)
- $\mathcal{G}_c = \emptyset$.

q	Performance	Std. Deviation
1	93.78%	2.92%
2	62.21%	7.84%
3	48.15%	9.71%
4	45.35%	11.11%
5	44.31%	11.79%

Effects of varying $\mathcal{G}_I^q$

Formulation (continuous action space)

- Multi-agent system with n agents; agent i picks actions $u_i \in R$.

- Agent i receives/measures private utility $f_i(\mathrm{u})$, where $u = \{u_1, \dots, u_n\}$.

- No models for the $f_i(\cdot)$.

- If collective action at time t is $u(t)$, agent i can only measure the numerical value $f_i(u(t))$.

- Collaborative objective – Welfare Optimization:

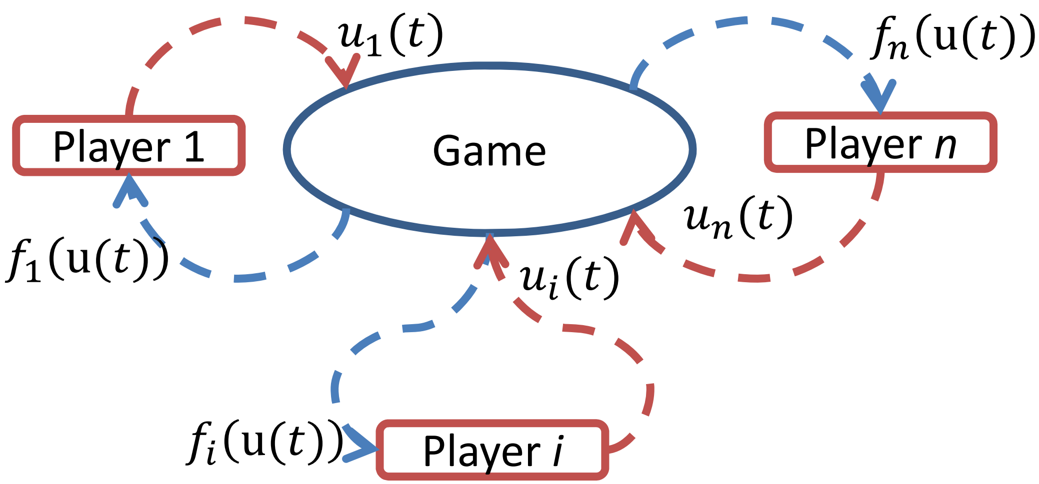
$$\min_{u \in R^n} W(u),$$

where $W(u) = \Sigma_{i=1}^{n} f_i(u)$.

Literature Review

→ Model-based distributed optimization techniques not applicable

→ Literature on Learning in Games is relevant.



Adaptation Loops of Players Playing a Repeated Game

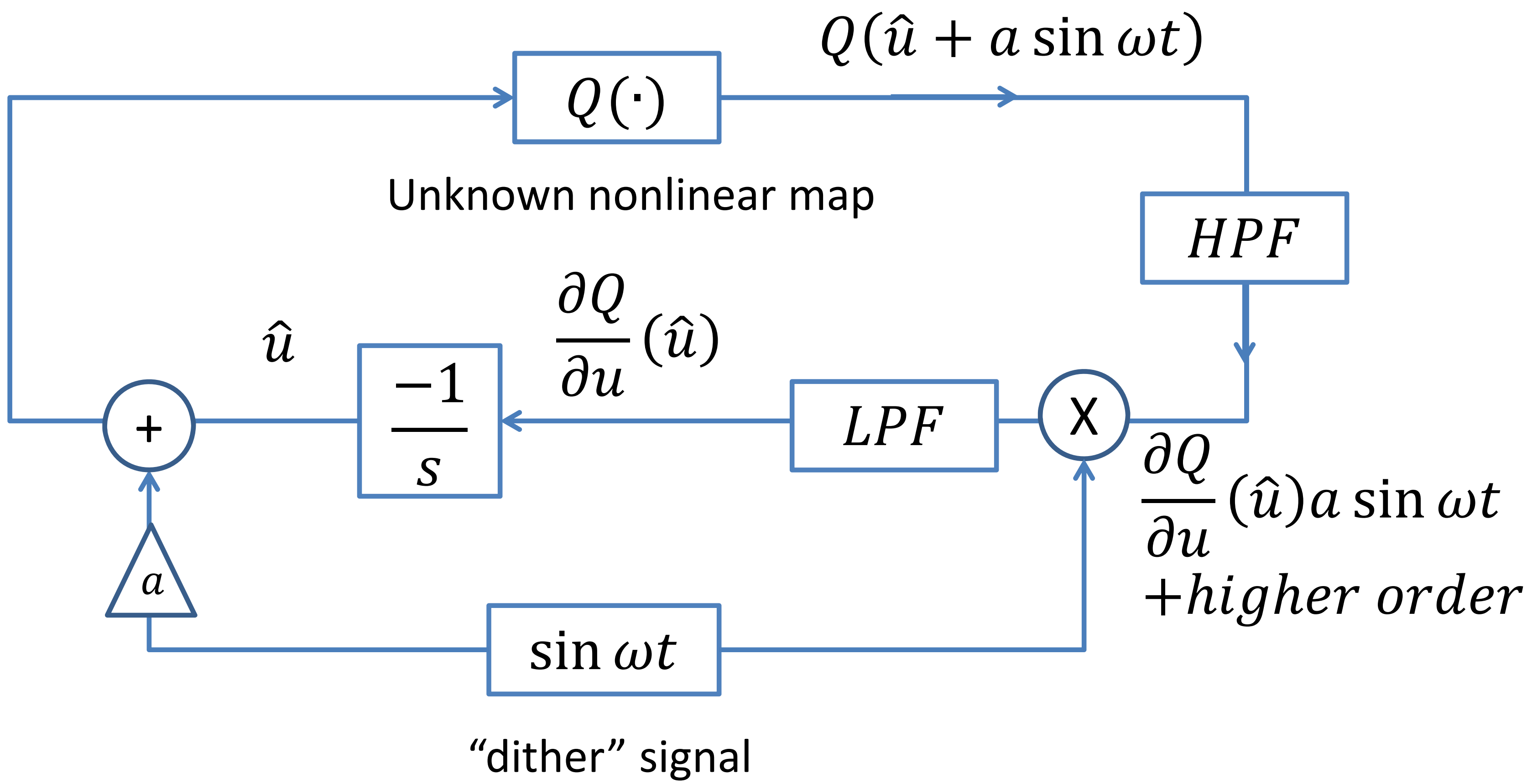
- Recent works [a,b] solve the problem using such ideas. But with discrete action sets – does not use gradient information → slow convergence.

- Recent works [c,d] use ideas from **extremum seeking control** for Nash seeking.

→ We go beyond Nash equilibration and use extremum seeking based ideas to achieve fast convergence to welfare optimal actions in this model-free setting.

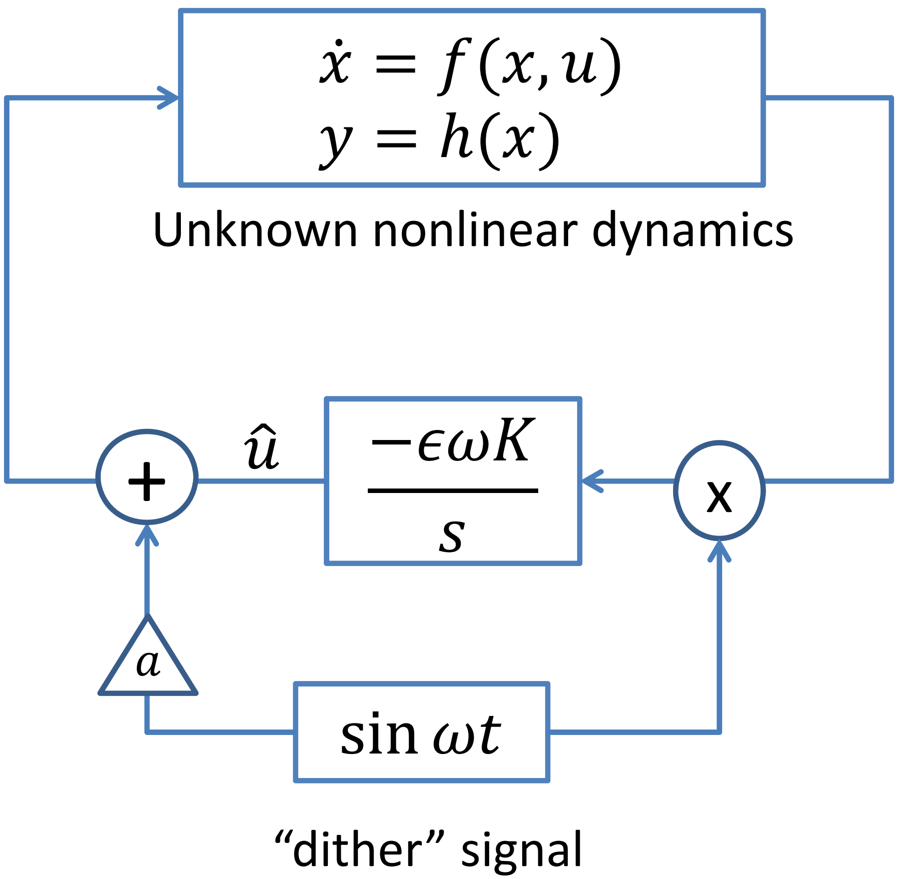
[a]. Marden, Young, Pao, "Achieving Pareto optimality through distributed learning," *IEEE CDC, 2012*.
[b]. Menon, Baras, "A distributed learning algorithm with bit-valued communications for multi-agent welfare optimization", IEEE CDC, 2013.
[c]. Frihauf, Krstic, Basar, "Nash equilibrium seeking in noncooperative games," IEEE Transactions on Automatic Control, 2012.
[d]. Stankovic, Johansson, Stipanovic, "Distributed seeking of Nash equilibria with applications to mobile sensor networks," IEEE TAC, 2012.

Extremum Seeking Control: Heuristics



$$\dot{\hat{u}} \approx -\frac{\partial Q}{\partial u}(\hat{u})$$

Extremum Seeking Control



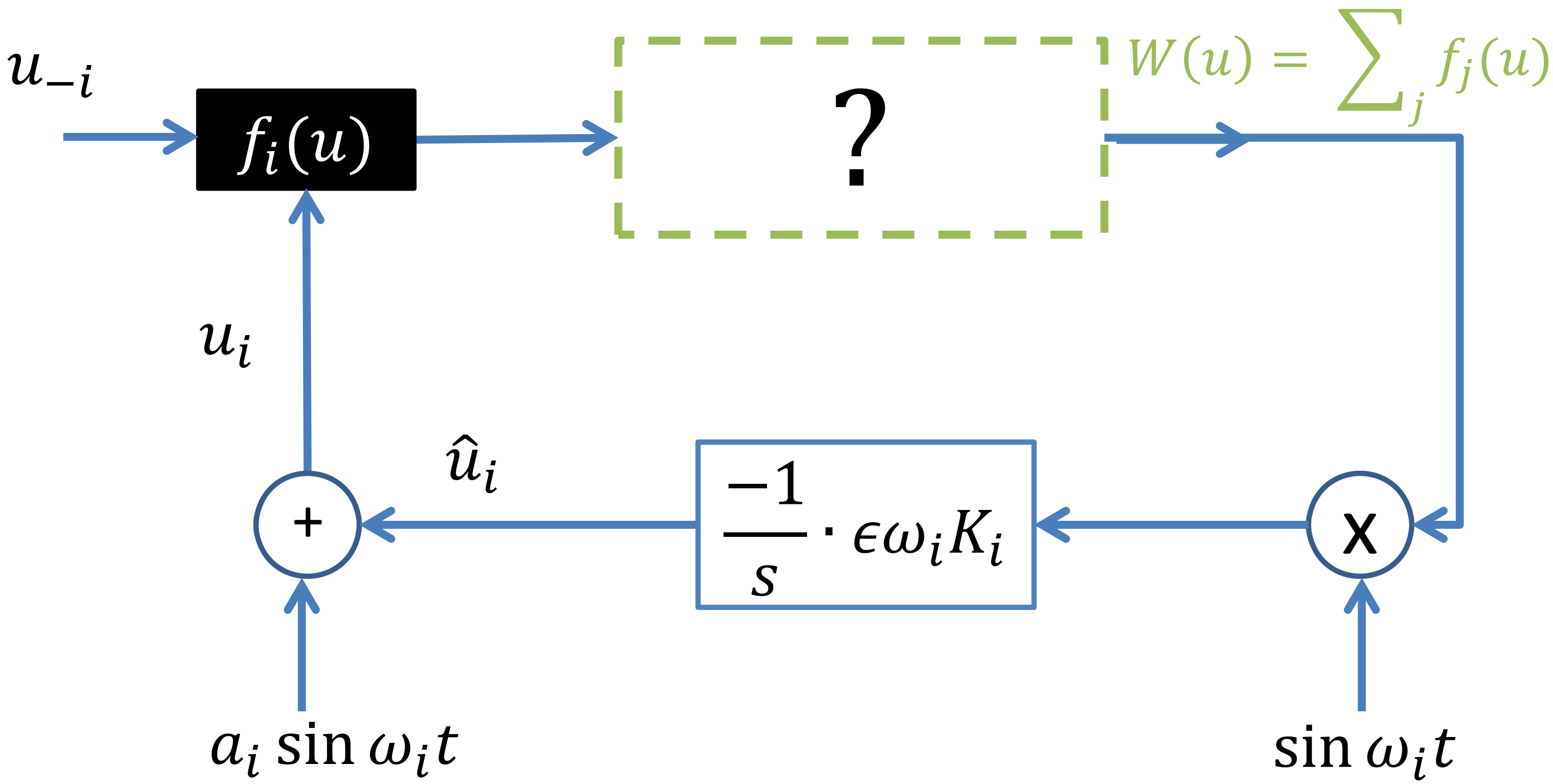
- Assuming there is an exponentially stable equilibrium $x^{eq} = l(u)$, for each u, the minimum of $h \circ l(\cdot)$ can be sought.

- Formal analysis uses singular perturbation and averaging arguments to prove local convergence of $\hat{u}$ to an $O(a + \omega + \epsilon)$ neighborhood of u^*.[a]

[a]. Krsti´c and Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, 2000.

Seeking the Welfare Optimal

Find a dynamical system that performs distributed averaging of reference signals of each agent.



Then, $\dot{\hat{u}}_i \approx -\frac{\partial W(u)}{\partial u_i}$.

Revisiting Dynamic Consensus

$(P0)$ Consider $\min_{\hat{x}} \sum_{i=1}^{n} (\hat{x} - r_i)^2$.

Taking derivative and setting it to zero → $\hat{x}^* = \frac{1}{n}\sum_{i=1}^{n} r_i$.

Now, consider the following reformulation of $(P0)$:

$(P1)$
$$\min \sum_{i=1}^{n} (x_i - r_i)^2 \;\; s.t. \, x_i = x_j \,, \qquad \forall\, i, j.$$

And finally, the following reformulation of $(P1)$:

$(P2)$
$$\min \frac{1}{2} x^T x - r^T x + \frac{1}{2} r^T r + \frac{1}{2} \rho x^T L_p x \,, \qquad s.t. \;\; L_I x = 0,$$

where L_I, L_P are graph Laplacians such that

$$Lx = 0 \Leftrightarrow x = \alpha \mathbf{1}.$$

→ The optimizer **doesn't change**: $\mathrm{x}^* = \frac{1}{n}\sum_{i=1}^{n} r_i \cdot \mathbf{1}$.

Revisiting Dynamic Consensus

$$(P2)\ \min \frac{1}{2}x^T x - r^T x + \frac{1}{2} r^T r + \frac{1}{2}\rho x^T L_p x, \qquad s.t.\ L_I x = 0.$$

Lagrangian for $(P2)$: $\mathcal{L}(x,\lambda) = \frac{1}{2}x^T x - r^T x + \frac{1}{2}\rho x^T L_p x + \lambda^T L_I x$

Optimal to $(P2)$ corresponds to a saddle point (x^*, λ^*):

$$\max_{\lambda} \mathcal{L}(x^*,\lambda) \le \mathcal{L}(x^*,\lambda^*) \le \min_{x} \mathcal{L}(x,\lambda^*).$$

So, consider the **saddle-seeking system**:

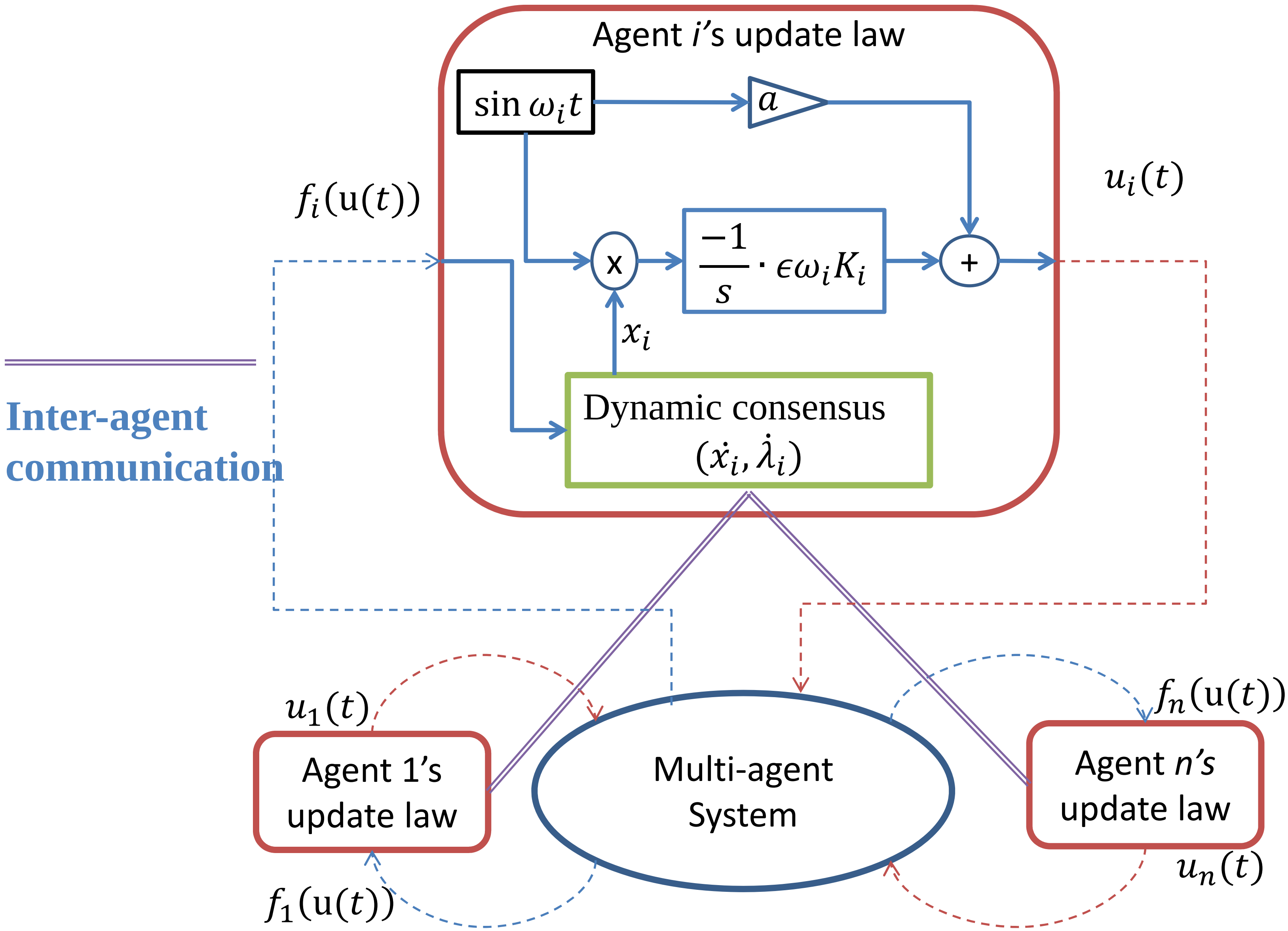
$$\begin{bmatrix}\dot{x}\\ \dot{\lambda}\end{bmatrix} = \begin{bmatrix}-\nabla_\lambda \mathcal{L}(x,\lambda)\\ \nabla_x \mathcal{L}(x,\lambda)\end{bmatrix} = \begin{bmatrix}-I - \rho L_P & -L_I^T\\ L_I & 0\end{bmatrix}\begin{bmatrix}x\\ \lambda\end{bmatrix} + \begin{bmatrix}I\\ 0\end{bmatrix} \mathrm{r}. \quad (1)$$

It can be proved this LTI system is stable, and its equilibrium verifies KKT conditions for $(P2)$. So

$$x(t) \to x^* = \frac{1}{n}\sum_{i=1}^{n} r_i \cdot \mathbf{1}.$$

While this algorithm has appeared in the literature earlier, our analysis is novel and essential for the formal proofs.

Proposed Solution



Proposed Solution Details

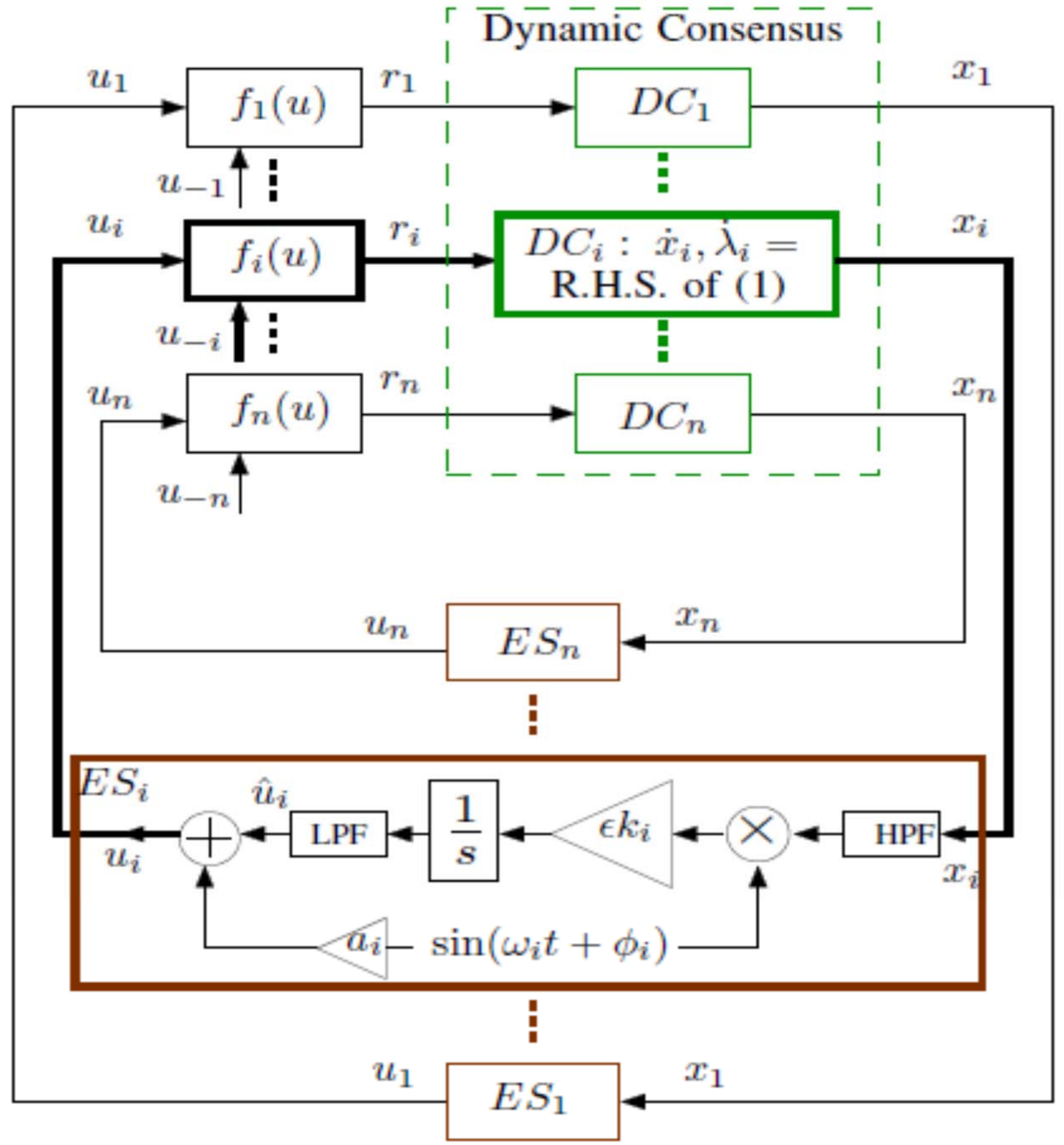


Fig. 1. A schematic representation of the proposed solution. DC_i refers to part of the dynamic consensus algorithm (1) implemented by agent i, ES_i refers to the extremum seeking law implemented by agent i, and u_{-i} refers to the elements of the vector u other than u_i.

Main Results[a]

Theorem [Dynamic Average Consensus (DAC)]: Let the undirected communication graph be connected, $\text{rank}(L_I) = (n-1)$, and $\frac{1}{2}\rho\lambda_{min}(L_P^T + L_P) < 1$. For a fixed $r(t) \equiv r$, the state of the DAC algorithm remains bounded and $x(t) \to \frac{1}{n}\sum_{i=1}^{n} r_i \cdot \mathbf{1}$ exponentially.

Theorem [Collaborative Welfare Seeking]: Let hypothesis of above Theorem hold, f_i be smooth, $\exists\, u^*\ s.t. \frac{\partial W(u^*)}{\partial u} = 0\,, \frac{\partial^2 W(u^*)}{\partial u^2} > 0$, and $\omega_i \neq \omega_j, 2\omega_i \neq \omega_k$, and $\omega_i \neq \omega_j + \omega_k$ for distinct i, j, k. Then there exists (ω, a, ϵ) small enough so that $u(t)$ converges to an $O(\|\omega\| + \epsilon + a)$ neighborhood of u^*, provided $\hat{u}(0)$ is sufficiently close to u^*.

The proof to the latter is based on averaging and singular perturbation arguments that are standard techniques in extremum seeking control theory.

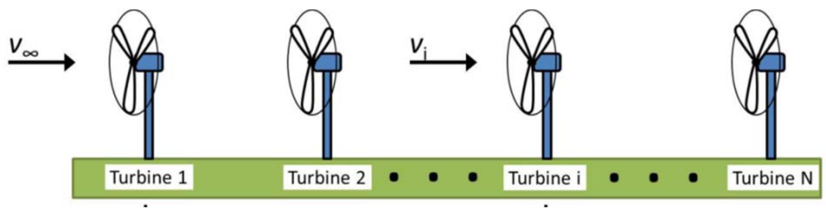
[a]. Menon, Baras, "Collaborative Extremum Seeking for Welfare Optimization", Submitted to *IEEE CDC, 2014*.

Wind Farm Power Maximization

Test model-free solution by simulating it on a wind farm model.

Wind Farm Model –

- Three turbines $n = 3$



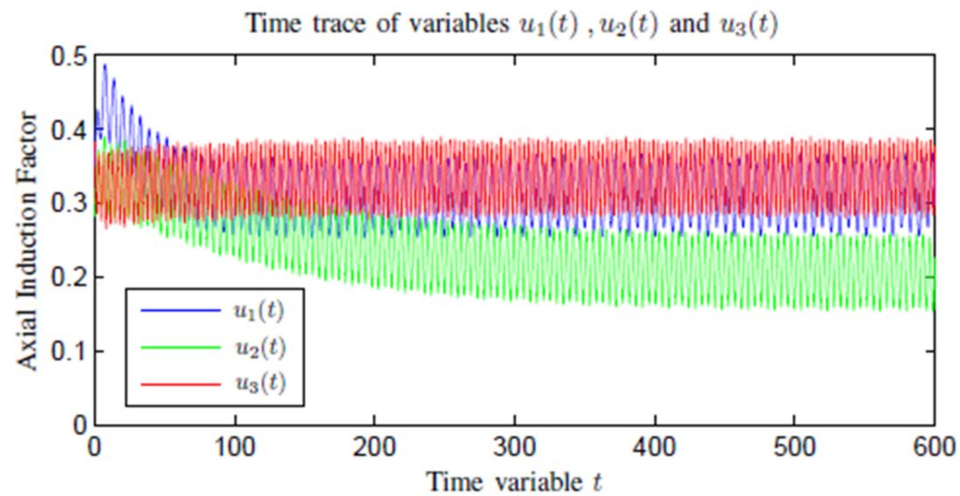
- Turbine action u_i is its Axial Induction Factor, $u_i \in [0,1/2]$.

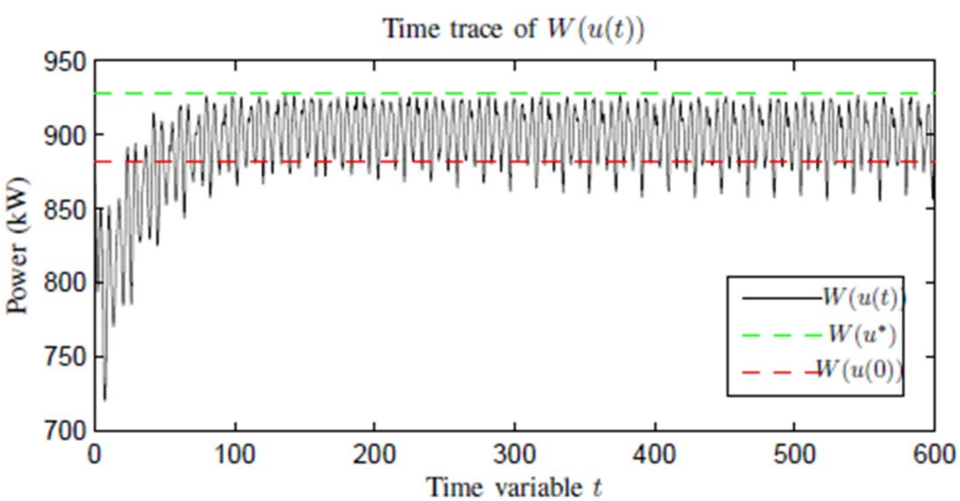
- Turbine power $f_i(u) = \frac{1}{2}\rho A_i C_p(u_i) V_i(u)^3$;
  - Where $C_p(u_i) = u_i\,(1 - u_i)^2$
  - $V_i(u)$ is the wind speed at turbine i, and is the coupling term
- Wake model

$$V_i(u) = V_\infty\left(1 - \sqrt{\sum_{j \in upstream(i)} \left(C[j,i]u_j\right)^2}\right)$$

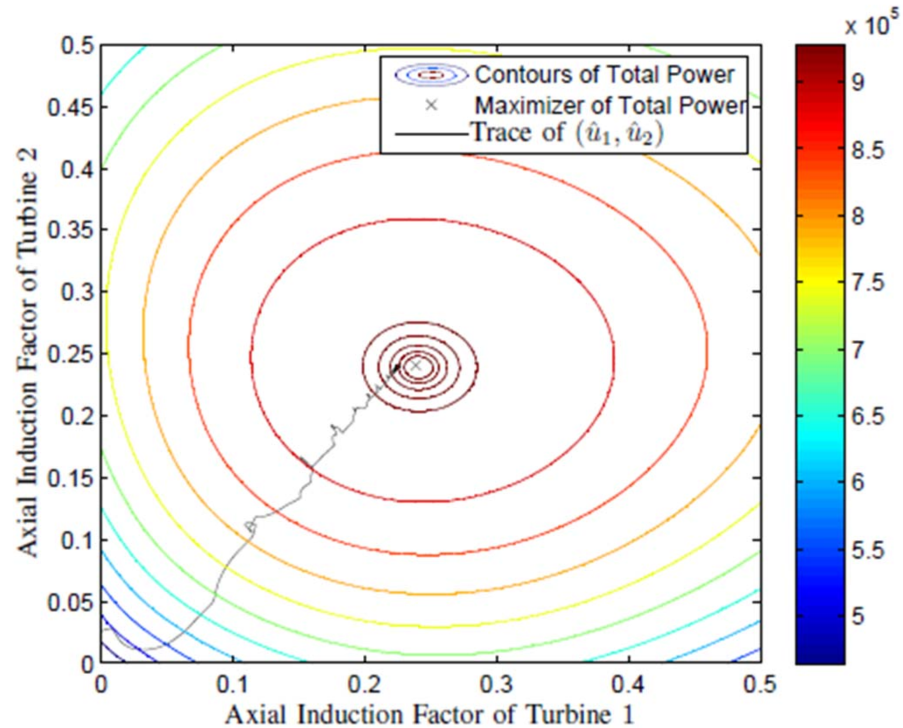
where the matrix C is computed based on the layout of the turbines (using the Park Model).

Wind Farm Simulation Results



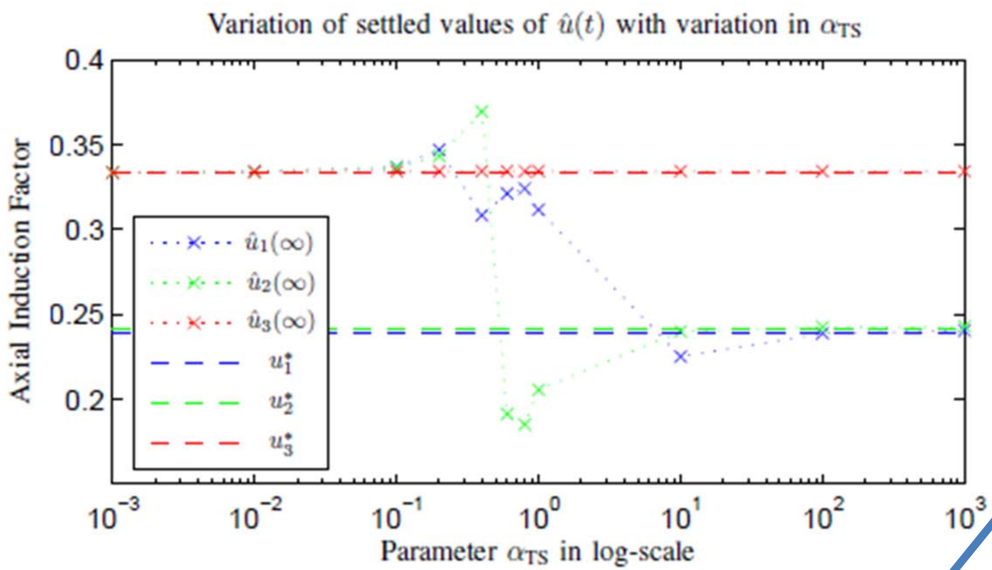


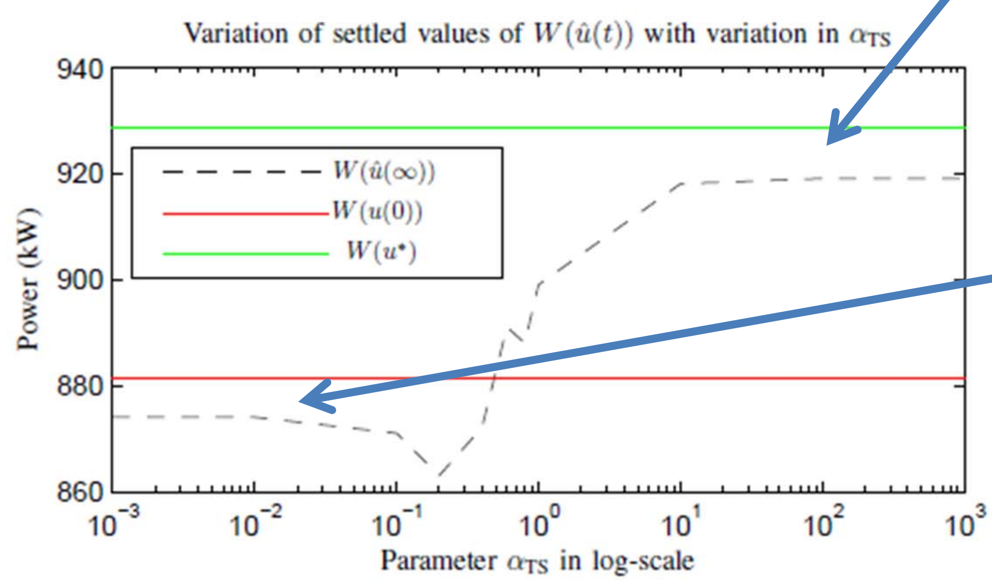
Typical run of the algorithm: Oscillations are expected in ESC due to additive injection of dither to u_i.



The “learning variable” $\hat{u}$ converges to a small neighborhood of u^*.

Learning vs. Consensus Time Scales





- The variable α_{TS} models the relative speed of the dynamic consensus and the learning dynamics.
- As long as the consensus is an order of magnitude faster than the learning dynamics, learning is successful.
- Else, $\hat{u}$ converges to a neighborhood of the "Nash equilibrium" where turbines optimize individual power.

Conclusions and Future Work

- Agents are influenced by their knowledge about the other agents' behavior in taking coordination decisions
- We modeled decision making on cooperation in a group effort as a result of two-person games on a network
- We studied adaptation to neighbors' strategies as a coordination mechanism using a learning algorithm
- The system is analyzed under classes of linear and bounded linear behavior functions. A generalized consensus problem determines strategy coordination
- The emerging collaboration graph is a function of agents' behavioral tendencies as well as the connectivity graph
- Exact results for complete graph developed. Future work will include extensions to other topologies.

Conclusions and Future Work

We demonstrated a distributed algorithm for multi-agent systems that

- exploits implicit and explicit communications
- to converge to welfare optimal actions
- without any model information.

Next steps

- speed of convergence?
- its dependence on $\mathcal{G}_c, \mathcal{G}_I$?
- continuous space analogs – general nonlinear systems – using gradient-type information for faster convergence?

Future Work

- Agents with general nonlinear dynamics
- Discrete time analog
- Effects of time-varying communication graph and structure of communication graph on the performance
- Application to collaborative robotics
- Detailed simulations on higher-fidelity wind farm models

References

[1] J.S. Baras and P. Hovareshti, “Efficient Communication Infrastructures for Distributed Control and Decision Making in Networked Stochastic Systems,” *Proceedings 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2010),* Budapest, Hungary, July 2010.

[2] J.S. Baras, P. Hovareshti and H. Chen, “Motif-based Communication Network Formation for Task Specific Collaboration in Complex Environments”, *Proceedings 2011 American Control Conference,* pp. 1051-1056, San Francisco, CA, June 2011.

[3] P. Hovareshti and J.S. Baras, “Learning Behaviors for Coordination in Networked Systems and a Generalized Consensus Protocol”, *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC 2011)*, pp. 2447-2452, Orlando, USA, December 2011.

[4] A. Menon and J.S. Baras, “Convergence Guarantees for a Decentralized Algorithm Achieving Pareto Optimality,” *Proceedings 2013 American Control Conference*, pp. 1935-1940, Washington, DC, June 2013

[5] A. Menon and J.S. Baras, "A Distributed Learning Algorithm with Bit-valued Communications for Multi-agent Welfare Optimization", *Proceedings 52nd IEEE Conference on Decision and Control*, pp. 2406-2411, Firenze, Italy, December 2013.

[6] A. Menon and J.S. Baras, “Collaborative Extremum Seeking for Welfare Optimization,” to appear in *Proceedings of the 53rd IEEE Conference on Decision and Control (CDC 2014*), Los Angeles, CA, December, 2014.

Thank you!

baras@umd.edu

http://www.isr.umd.edu/~baras

Questions?