Attitude coordination with potential game approach by learning algorithm SAP

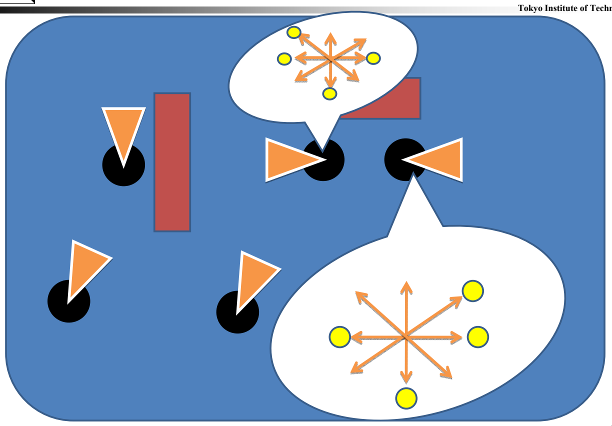
Goto Tatsuhiko
FL09_19_1

14th,January,2010

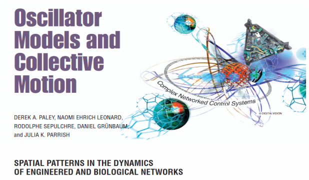
Outline

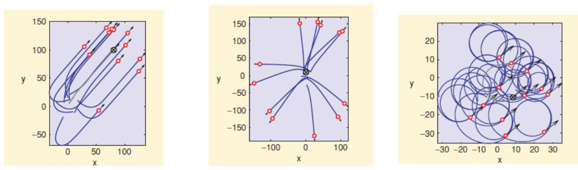
- Background
- Outline of Oscillator
- Review of Potential game (RSAP)
- Adapt Potential game to Oscillator Problem
- Simulation Result
- Experimental Result
- Summary and assignment

Background



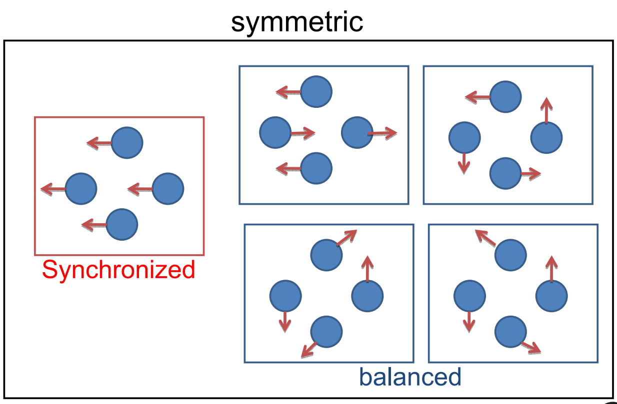
Outline of Oscillator





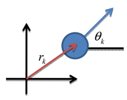
synchronized | balanced | Synchronized with natural frequency

Synchronized, balanced



Oscillator

Particle model



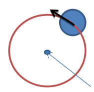
$r_k = x_k + jy_k$

$\dot{r}_k = e^{i\theta_k}$

$\dot{\theta}_k = u_k(r,\theta)$

$u_k = 0$

$u_k = \omega_0$



$c_k = r_k + \omega_0^{-1} i e^{i\theta_k}$

Control law

$$u_k = \omega_o + u_k^{spac}(r,\theta) + u_k^{ori}(\theta)$$

Natural frequency — Spacing control — Orientation control

Not consider the position

$$u_k = \omega_o + u_k^{ori}(\theta)$$

Depend only on relative phase

$\theta_{kj} = \theta_k - \theta_j$

Oscillator

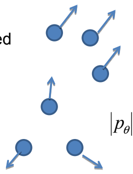
$\theta_k = \theta_j$ Agent j and k is parallel position

Phase order parameter

$$p_\theta = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j} \qquad 0 \le |p_\theta| \le 1$$

$|p_\theta| = 1$ ➡ synchronized

$|p_\theta| = 0$ ➡ balanced



$$|p_\theta| = \left|\frac{1}{3}(e^{\pi/4}\times 3)\right| = |e^{\pi/4}| = 1$$

$$|p_\theta| = \left|\frac{1}{3}(e^{\pi/2} + e^{7\pi/6} + e^{11\pi/6})\right| = 0$$

$p_\theta = \dot{R}$ $\quad p_\theta$ is equal to the velocity of R $\quad R \triangleq \frac{1}{N}\sum_{j=1}^{N} r_j$. Center of mass

Oscillator

Phase potential

$$U_1(\theta) = \frac{N}{2}|p_\theta|^2$$

$|p_\theta| = 1$ ➡ synchronized
$|p_\theta| = 0$ ➡ balanced
➡
U_1 maximize ➡ synchronized
U_1 minimize ➡ balanced

Gradient control

$$\frac{\partial U_1}{\partial \theta_k} = \langle ie^{i\theta_k}, p_\theta \rangle$$ ➡

$$u_k = \omega_o - k_1 \langle ie^{i\theta_k}, p_\theta \rangle$$
$K_1 < 0 \Rightarrow synchronized$
$K_1 > 0 \Rightarrow balanced$

$\langle x, y \rangle = \mathrm{Re}\{x^* y\}$

Use all agents information

➡ $u_k = \omega_o + \frac{k_1}{N}\sum_{j=1}^{N} sin\theta_{kj}$

Kuramoto model

Oscillator

Phase potential $\quad U_1(\theta) = \frac{N}{2}|p_\theta|^2$ $\quad$ Use all agents information

⬇

Laplacian Phase Potential

$$W_1(\theta) = Q_L(e^{i\theta}) = \frac{1}{2N}\langle e^{i\theta}, Le^{i\theta} \rangle$$

Laplacian quadratic form

$$Q_L(x) = \frac{1}{2N}\langle x, Lx \rangle$$
$$\langle x, y \rangle = \sum_{j=1}^{N} \langle x_j, y_j \rangle$$

Interaction is limited is OK

Graph is complete

minimize	Synchronized phase
maximize	Balanced phase

$$\rightarrow W_1(\theta) = \frac{N}{2} - U_1(\theta)$$

$$W_1(\theta) = Q_L(e^{i\theta}) = \frac{1}{2N}\langle e^{i\theta}, Le^{i\theta} \rangle$$

$$= \frac{1}{2N}\sum_{k=1}^{N}(d_k - \sum_{j \in N_k}\langle e^{i\theta_k}, e^{i\theta_j} \rangle) \ge 0$$

$$W_1(\theta) = 0 \Leftrightarrow synchrozized$$

Oscillator

$$W_1(\theta) = Q_L(e^{i\theta}) = \frac{1}{2N}\langle e^{i\theta}, Le^{i\theta} \rangle$$

Laplacian Phase control (continuous)

$$\frac{\partial W_1(\theta)}{\partial \theta_k} = \frac{1}{2N}\langle ie^{i\theta_k}, (L + L^T)_k e^{i\theta} \rangle = \frac{1}{N}\langle ie^{i\theta_k}, L_k e^{i\theta} \rangle$$

$$u_k^{ori} = \frac{k_1}{N}\langle ie^{i\theta_k}, L_k e^{i\theta} \rangle \text{ (gradient control)} \qquad \dot{\theta}_k = \omega_0 + u_k^{ori}$$

⬇

Simplified Kuramoto model $\quad \dot{\theta}_k = \omega_0 + \frac{K_1}{N}\sum_{j \in N_k} \sin\theta_{kj}$

➡ Evolves monotonically

$$\dot{W}_1(\theta) = \frac{k_1}{N^2}\sum_{k=1}^{N}\langle e^{i\theta_k}, L_k e^{i\theta} \rangle^2$$

Invariance principle

Converge to largest invariant set for which $\dot{W}_1(\theta) = 0$

Oscillator

$\dot{W}_1(\theta) = 0$ ➡ $\langle ie^{i\theta_k}, L_k e^{i\theta} \rangle = 0$ $\qquad \dot{W}_1(\theta) = \frac{k_1}{N^2}\sum_{k=1}^{N}\langle e^{i\theta_k}, L_k e^{i\theta} \rangle^2$

⬌ θ is stationary point
(ex) $e^{i\theta}$ is an eigenvector of L

1^T is an eigenvector of L (from graph theorem)
$e^{i\theta} = 1^T$ synchronized $\quad K_1 < 0$

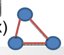
➡

graph is circulant
⬇
$e^{i\theta}(1^T e^{i\theta} = 0)$ is an eigenvector of L (from graph theorem)
$1^T e^{i\theta} = 0$ balanced $\quad K_1 > 0$

circulant

$$L = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{n-2} & & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{bmatrix}$$

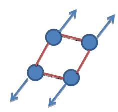
(ex)



Oscillator

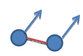
Pi/2 problem

$$\dot{\theta}_k = \omega_0 + \frac{K_1}{N}\sum_{j \in N_k} \sin\theta_{kj}$$



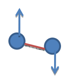
If $\sin\theta_{kj} = 0$,then $\dot{\theta}_k = \omega_0$

When realizing synchronized



⬇

$$\sin(\pi/2 - (-\pi/2)) = 0$$



Evolution stop but not realize synchronized

- Background
- Outline of Oscillator
- Review of Potential game (RSAP)
- Adapt Potential game to Oscillator Problem
- Simulation Result
- Experimental Result
- Summary and assignment

Global planner $\phi : A \rightarrow R$ (potential function)

aligned

Player's objective function

$$U_i(a_i'', a_{-i}) - U_i(a_i', a_{-i}) = \phi_i(a_i'', a_{-i}) - \phi_i(a_i', a_{-i})$$

→ Changing in the player's objective function = Changing in the potential function

Consensus point

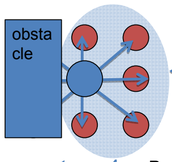
→ Nash equilibrium of the game characterized by the Player's objective function $U_i(a) = -\sum_{p_j \in N_i} \|a_i - a_j\|$

→ set of Consensus point $A^c \subset A^*$ set of Nash equilibrium

Learning algorithm

Guarantees probabilistic convergence to a pure Nash equilibrium that maximize the potential function

Issue of SAP → Permit any action in its action set



→ From mobility limitation , it's impossible

← Restricted action set $R_i(a_i(t-1))$

Restrictive Spatial Adaptive Play

step1 Randomly choose one player P_i

(Another player do a same action)

step2 Player P_i selects **one trial action** $\hat{a}_i$ **from** $R_i(a_i(t-1))$. $z_i = \max_{a_i \in A_i} |R_i(a_i)|$

$$\Pr[\hat{a}_i = a_i] = 1/z_i \quad , \quad a_i \in R_i(a_i(t-1)) \setminus a_i(t-1)$$

$$\Pr[\hat{a}_i = a_i(t-1)] = 1 - (|R_i(a_i(t-1))| - 1)/z_i$$

step3 Player P_i chooses its action at time t

$$\Pr[a_i(t) = \hat{a}_i] = \frac{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\}}{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\} + \exp\{\beta U_i(a(t-1))\}}$$

→ Choose trial action

$$\Pr[a_i(t) = a_i(t-1)] = \frac{\exp\{\beta U_i(a(t-1))\}}{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\} + \exp\{\beta U_i(a(t-1))\}}$$

→ Don't move

Theorem

Consider a finite n-player potential game with potential function ϕ

If restricted action set satisfy

$a_i^2 \in R_i(a_i^1) \leftrightarrow a_i^1 \in R_i(a_i^2)$ (Reversibility)

and $a_i^k \in R_i(a_i^{k-1})$ (Feasibility)

then RSAP induce a Markov process over the state space A where the unique distribution $\mu \in \Delta(A)$ is

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in A} \exp\{\beta\phi(\bar{a})\}}$$

Outline

- Background
- Outline of Oscillator
- Review of Potential game (RSAP)
- Adapt Potential game to Oscillator Problem
- Simulation Result
- Experimental Result
- Summary and assignment

Synchronized

$$W_1(\theta)=\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\langle e^{i\theta_k},e^{i\theta_j}\rangle)$$
$$=\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\cos(\theta_k-\theta_j))$$

→ minimize

Synchronized phase

Potential game

$$\phi(\theta)=-\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\cos(\theta_k-\theta_j))$$

$$U_i(\theta)=\frac{1}{N}\sum_{j\in N_i}\cos(\theta_i-\theta_j)$$

(Time-invariant interaction graph and undirected)

(proof) $$\phi(\theta)=-\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\cos(\theta_k-\theta_j))$$
$$=-\frac{1}{2N}(\sum_{k=1}^{N}d_k-\sum_{j\in N_i}2\cos(\theta_i-\theta_j)-\sum_{j\neq i}\sum_{k\in N_j\setminus i}\cos(\theta_j-\theta_k))$$

Interaction fixed

$$\phi(\theta_i^{''},\theta_{-i})-\phi(\theta_i^{'},\theta_{-i})=\frac{1}{2N}\sum_{j\in N_i}2\cos(\theta_i^{''}-\theta_j)-\frac{1}{2N}\sum_{j\in N_i}2\cos(\theta_i^{'}-\theta_j)$$
$$=U_i(\theta_i^{''},\theta_{-i})-U_i(\theta_i^{'},\theta_{-i})$$

Balanced

$$W_1(\theta)=\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\langle e^{i\theta_k},e^{i\theta_j}\rangle)$$
$$=\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\cos(\theta_k-\theta_j))$$

→ maximize

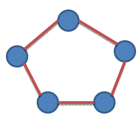
Balanced phase

Potential game

$$\phi(\theta)=\frac{1}{2N}\sum_{k=1}^{N}(d_k-\sum_{j\in N_k}\cos(\theta_k-\theta_j))$$

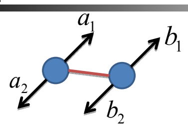
$$U_i(\theta)=-\frac{1}{N}\sum_{j\in N_i}\cos(\theta_i-\theta_j)$$

(Time-invariant interaction graph and undirected)



Graph is circulant

example



2 agents
2 actions → 4 states

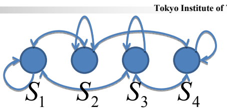
$S_1=(a_1,b_1)$
$S_2=(a_1,b_2)$
$S_3=(a_2,b_1)$
$S_4=(a_2,b_2)$

Transmission matrix

$$P=\begin{bmatrix} 1-p_{12}-p_{13} & \frac{1}{4}\frac{e^{\beta U_2(S_2)}}{e^{\beta U_2(S_1)}+e^{\beta U_2(S_2)}} & \frac{1}{4}\frac{e^{\beta U_1(S_3)}}{e^{\beta U_1(S_1)}+e^{\beta U_1(S_3)}} & 0 \\ \frac{1}{4}\frac{e^{\beta U_2(S_1)}}{e^{\beta U_2(S_1)}+e^{\beta U_2(S_2)}} & 1-p_{21}-p_{24} & 0 & \frac{1}{4}\frac{e^{\beta U_1(S_4)}}{e^{\beta U_1(S_4)}+e^{\beta U_1(S_2)}} \\ \frac{1}{4}\frac{e^{\beta U_1(S_1)}}{e^{\beta U_1(S_1)}+e^{\beta U_1(S_3)}} & 0 & 1-p_{31}-p_{34} & \frac{1}{4}\frac{e^{\beta U_2(S_4)}}{e^{\beta U_2(S_3)}+e^{\beta U_2(S_4)}} \\ 0 & \frac{1}{4}\frac{e^{\beta U_1(S_2)}}{e^{\beta U_1(S_4)}+e^{\beta U_1(S_2)}} & \frac{1}{4}\frac{e^{\beta U_2(S_3)}}{e^{\beta U_2(S_3)}+e^{\beta U_2(S_4)}} & 1-p_{42}-p_{43} \end{bmatrix}$$

example

Transmission matrix P → irreducible, aperiodic



Exist Stationary distribution

$$\mu(a)=\frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a}\in A}\exp\{\beta\phi(\bar{a})\}}$$

β const $\quad \mu(a)p_{ab}=\mu(b)p_{ba}$ (detail balance)

(ex)

$\beta=0$ → $$P=\begin{bmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{3}{4} & 0 & \frac{1}{8} \\ \frac{1}{8} & 0 & \frac{3}{4} & \frac{1}{8} \\ 0 & \frac{1}{8} & \frac{1}{8} & \frac{3}{4} \end{bmatrix}$$

Stationary distribution

$$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

example

(ex)

$\beta=\infty$ → $$P=\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

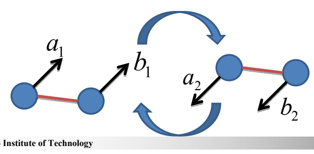
Absorbed state

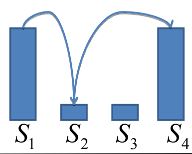
↓

Not exist stationary distribution

SAP's features is not convergence one state!

If there are some consensus point, change the state during some consensus point.



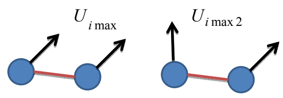


β 's degree

What degree is better for β ?

The most largest probability from consensus point to another state is

$$\max_{i \in N} \frac{1}{nz} \frac{e^{\beta U_{i\max 2}}}{e^{\beta U_{i\max}} + e^{\beta U_{i\max 2}}}$$



Probability of changing the state from Consensus point is smaller than

$$\sum_{i \in N} (z_i^{'} - 1) \times \max_{i \in N} \frac{1}{nz} \frac{e^{\beta U_{i\max 2}}}{e^{\beta U_{i\max}} + e^{\beta U_{i\max 2}}}$$

$z_i^{'}$:feasible action of i

β 's degree

spec

Do small the probability of changing the state from Consensus point than p

$$\sum_{i \in N} (z_i^{'} - 1) \times \max_{i \in N} \frac{1}{nz} \frac{e^{\beta U_{i\max 2}}}{e^{\beta U_{i\max}} + e^{\beta U_{i\max 2}}} < p$$

$$\max_{i \in N} \frac{1}{U_{i\max} - U_{i\max 2}} \log \left(\frac{\sum_{i \in N} (z_i^{'} - 1)}{pnz} - 1 \right) < \beta$$

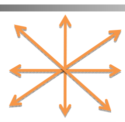
A little strict condition

Outline

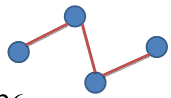
- Background
- Outline of Oscillator
- Review of Potential game (RSAP)
- Adapt Potential game to Oscillator Problem
- Simulation Result
- Experimental Result
- Summary and assignment

Simulation Result (synchronized non obstacle)

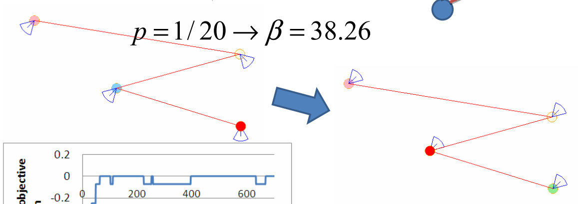
8 action sets

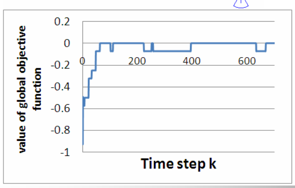


Interaction graph



$p = 1/20 \rightarrow \beta = 38.26$

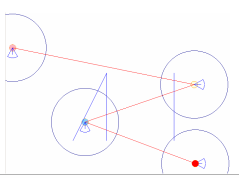




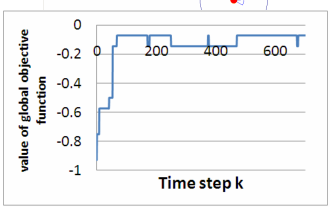
One consensus point (k=500)

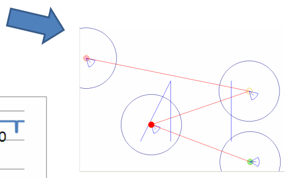
Simulation Result (synchronized with obstacle)

$p = 1/20 \rightarrow \beta = 29.0$



Circle is sensor area of obstacle

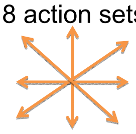




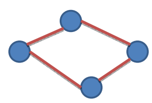
One consensus point (k=500)

Simulation Result (balanced with four agents)

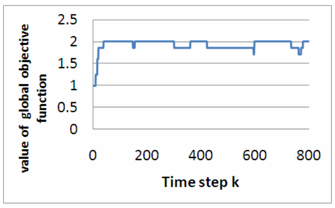
8 action sets



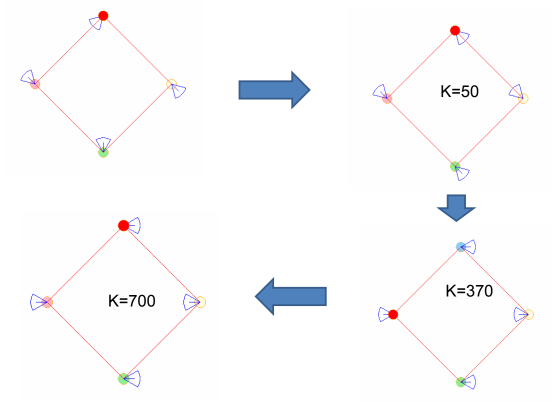
Interaction graph



$p = 1/20 \rightarrow \beta = 19.14$

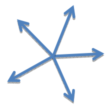


Simulation Result (balanced with four agents)

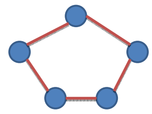


Simulation Result (balanced with five agents)

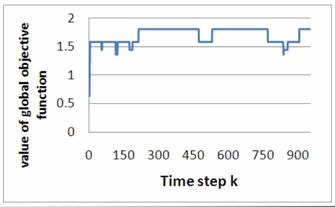
5 action sets



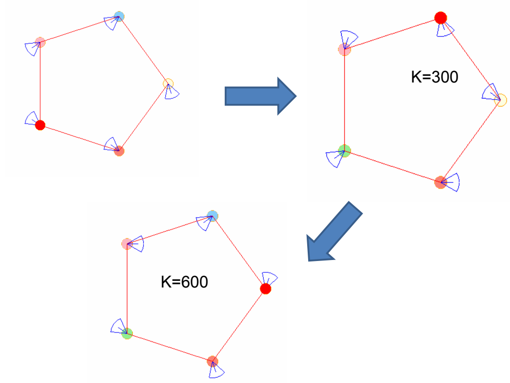
Interaction graph



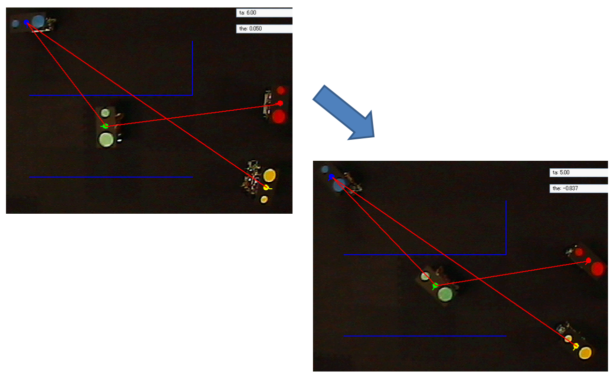
$p = 1/80 \rightarrow \beta = 18.2$



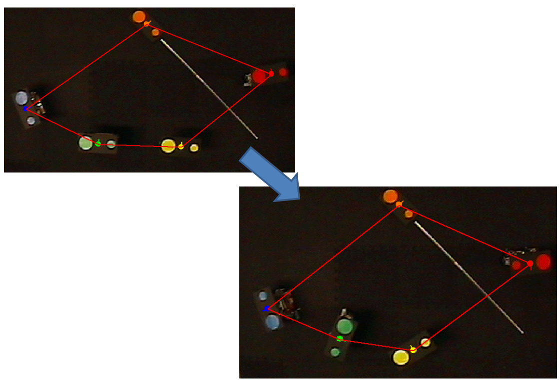
Simulation Result (balanced with five agents)



Experimental Result (synchronization)



Experimental Result (balanced)



Summary and assignment

- Can adapt potential game to oscillator
- RSAP can make Stationary distribution
- Some time change consensus state to another consensus point
- Can decide changing rate with exploration rate β
- How to make suboptimal state high probability
  →using reinforcement learning？
- How to decide p
  →consider not consensus states