2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

# Experiment on spatiotemporal dynamics generation and parallel decision making using spatial light modulator and semiconductor laser

Kento Takehana†, Kensei Morijiri†, Takatomo Mihana†‡,
Kazutaka Kanno†, Makoto Naruse‡, and Atsushi Uchida†

†Department of Information and Computer Sciences, Saitama University, Japan
‡Department of Information Physics and Computing, Graduate School of
Information Science and Technology, The University of Tokyo, Japan
Email: k.takehana.802@ms.saitama-u.ac.jp, auchida@mail.saitama-u.ac.jp

**Abstract–** We experimentally demonstrate parallel implementation of photonic decision making for solving the multi-armed bandit problem using a spatial light modulator, a camera, and a semiconductor laser. We achieve experimental decision making in a multi-armed bandit problem with up to 512 slot machines using chaotic spatiotemporal dynamics generated from the semiconductor laser.

## 1. Introduction

Reinforcement learning has been used in various research fields, and the multi-armed bandit problem [1] is an example of reinforcement learning. The goal of the multi-armed bandit problem is to maximize the total rewards in which a player repeatedly selects multiple slot machines with different unknown hit probabilities. To solve the multi-armed bandit problem, two actions of "exploration" and "exploitation" must be balanced, and the two actions are in a trade-off relationship. Exploration is an action to search for the slot machine with the highest hit probability by selecting multiple slot machines randomly. On the contrary, exploitation is an action to repeatedly play the slot machine with the highest hit probability estimated by the exploration to increase the total rewards.

Photonic decision making for solving the multi-armed bandit problem has been reported using chaotic semiconductor lasers [2-4] based on the tug-of-war method [5,6]. In particular, a method using multiple temporal waveforms of semiconductor lasers has been proposed to solve the multi-armed bandit problem with up to 1024 slot machines numerically [6]. However, this method requires the same number of semiconductor lasers as the number of slot machines. It is difficult to experimentally demonstrate decision making for a large number of slot machines using this method.

An experimental implementation of reservoir computing has been reported using a spatial light modulator [7,8]. A network with a large number of nodes can be realized by utilizing the parallelism of spatial light. It is expected to solve the multi-armed bandit problem with many slot machines using spatial light modulator.

In this study, we demonstrate the generation of spatiotemporal dynamics using an optoelectronic feedback system with a semiconductor laser, a camera, and a spatial light modulator. We perform decision making for solving the multi-armed bandit problem using this optoelectronic feedback system.

## 2. Methods

Our experimental setup is shown in Fig. 1. The laser beam emitted from a collimator becomes a spherical wave by passing through a spatial filter, and is injected into a spatial light modulator. The spatial light modulator modulates the phase of the incident light by the input voltage of the pixels on the spatial light modulator. Intensity modulation can be achieved by combining two polarizers and the spatial light modulator. The intensity-modulated laser output is captured by a CMOS camera and transmitted to a computer. The post processing of the detected image is performed in the computer and the image signal is fed back to the spatial light modulator. Then, the spatial light modulator modulates the phase of the laser light with a new input values. Thus, spatiotemporal dynamics can be generated by repeating the detection and feedback of the image in the optoelectronic feedback system.

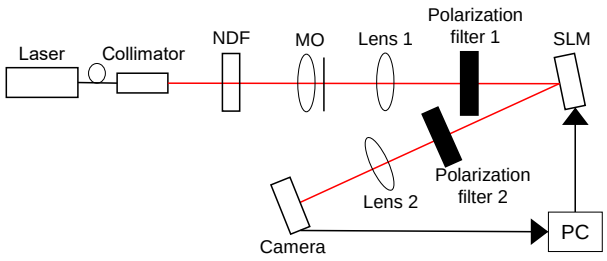


Fig. 1 Experimental setup for spatiotemporal dynamics using optoelectronic feedback system with semiconductor laser, camera, and spatial light modulator.

The intensity dynamics of this feedback system for each pixel can be can be modeled as follows.

$$I^{CAM} = a \cdot \cos(2\pi f I^{SLM}) + b \qquad (1)$$

where $I^{CAM}$ is the value of the macro pixel on the camera, $I^{SLM}$ is the value of the macro pixel on the spatial light

modulator. $a$ is the amplitude of the intensity modulation, $f$ is the frequency of the intensity modulation, and $b$ is the bias of the intensity modulation.

The feedback signal from the computer to the spatial light modulator can be described as follows.

$$I^{SLM}(t+1) = \beta \cdot I^{CAM}(t) \tag{2}$$

where $\beta$ is the feedback strength and it is the parameter that determines the spatiotemporal dynamics.

From Eqs. (1) and (2), one-dimensional map of the feedback system for each pixel can be written as follow.

$$I^{CAM}(t+1) = a \cdot \cos\left(2\pi f \beta I^{CAM}(t)\right) + b \tag{3}$$

Equation (3) shows that the spatiotemporal dynamics generated by the feedback system is based on a sinusoidal function

## 3. Generation of spatiotemporal dynamics

In this section, we observe spatiotemporal dynamics generated in the experimental setup. A bifurcation to chaos is observed by changing the feedback strength. Figure 5 shows the spatiotemporal dynamics at the feedback strength $\beta = 3.2$. Irregular spatiotemporal patterns are observed at the different number of iterations in Fig. 2.

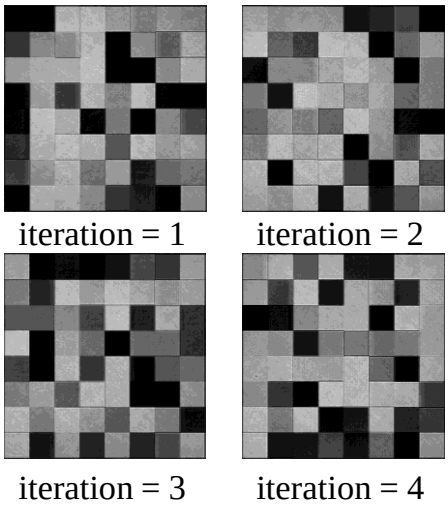


Fig. 2 Spatiotemporal patterns of spatial light modulator at four successive iterations at the feedback strength $\beta = 3.2$.

## 4. Experiment on decision making

### 4.1 Decision-making method

In this section, we describe the decision-making procedure. Figure 3 shows a scheme for decision making using spatiotemporal dynamics. Slot machines are assigned to macro pixels of the spatial light modulator. The temporal dynamics of macro pixels are used for slot machine selection. The bias is added to the temporal waveform of each macro pixel as follows.

$$F_i(t) = C_i(t) + kX_i(t) \tag{4}$$

where $C_i(t)$ is the chaotic temporal waveform of $i$-th macro pixel, $k$ is the bias coefficient, $X_i(t)$ is the bias of $i$-th macro pixel. We select the $i$-th slot machine corresponding to the maximum value of the temporal waveforms among the biased signal $F_i(t)$. The bias is updated based on the reward of the selected slot machine as follows.

$$X_i(t) = Q_i(t) - \frac{1}{n-1}\sum_{i' \neq i}^{n} Q_{i'}(t) \tag{5}$$

$$Q_i(t) = \Delta W_i - \omega L_i \tag{6}$$

$$\Delta = 2 - \left(\hat{P}_{top1} + \hat{P}_{top2}\right) \tag{7}$$

$$\omega = \hat{P}_{top1} + \hat{P}_{top2} \tag{8}$$

$$\hat{P} = \frac{W_i}{N_i} \tag{9}$$

where $N_i$ is the number of selection for the $i$-th slot machine, $W_i$ is the number of "hits" for the $i$-th slot machine, $L_i$ is the number of "miss" for the $i$-th slot machine, $\hat{P}$ is the estimated hit probability of the $i$-th slot machine. Decision making is achieved by iteratively selecting one of the slot machines and updating the bias.

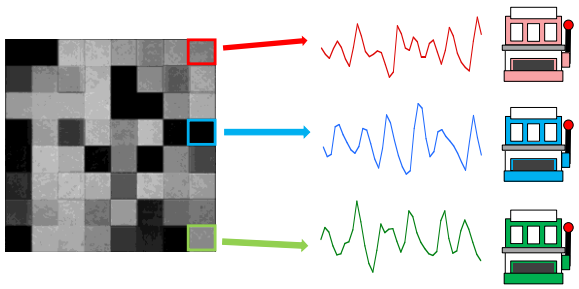

Fig. 3 Correspondence between spatiotemporal dynamics and slot machines.

Table 1 Parameter settings for multi-armed bandit problem

| | |
|---|---|
| Hit probability of slot machine 1 | 0.7 |
| Hit probability of slot machine 2 | 0.5 |
| Hit probability of slot machine 3 | 0.9 |
| Hit probability of slot machine 4 | 0.1 |
| Hit probability of slot machine $2m$+1 | 0.7 |
| Hit probability of slot machine $2m$+2 | 0.5 |
| Reward | 0 or 1 |
| Number of cycles | 128 |
| Number of plays per cycle | 10000 |
| Bias coefficient $k$ | 15 |

### 4.2 Parameter settings

The parameter values for the multi-armed bandit problem are summarized in Table 1. Here, $m$ is an integer ($m \geq 2$) and the bias coefficient $k$ is the parameter used in Eq. (4).

## 5. Results

### 5.1 Evaluation of decision making

In this study, the correct decision rate (CDR) and the regret are used for the quantitative evaluation of decision-making performance. CDR is the rate of the cycles in which the correct decision is made. When $n$ cycles of decision making are conducted with $m$ plays, CDR at the $t$-th cycle is defined as follows.

$$CDR(t) = \frac{1}{n}\sum_{i=1}^{n} C(i,t) \quad (1 \leq t \leq m) \qquad (10)$$

$$C(i,t) = \begin{cases} 1 & \text{(Selection of correct slot machine)} \\ 0 & \text{(otherwise)} \end{cases}$$

where $C(i,t)$ is the function that returns 1 if the selected slot machine is the correct slot machine, and 0 otherwise.

Regret is defined as the difference between the ideal total reward and the actual reward.

$$Regret(t) = t\,P_{max} - \frac{1}{S}\sum_{l=1}^{S}\sum_{i=1}^{I}\left(P_i\,N_{l,i}(t)\right) \qquad (11)$$

where $P_{max}$ is the maximum hit probability, $S$ is the total number of cycles, $P_i$ is the hit probability of the $i$-th slot machine, $N_{l,i}$ is the number of selection for the $i$-th slot machine up to the $t$-th play in the $l$-th cycle.

### 5.2 Experimental results

Figure 4(a) shows the results of CDR for 8, 64, and 512 slot machines. CDR is larger than 0.95 and correct decision-making is achieved up to 512 slot machines. Figure 4(b) shows the scalability of number of plays for $CDR = 0.95$ as the number of slot machines is changed. The graph is approximated by a power law. The exponent of 0.902 of the power law is obtained as the number of slot machine $N$ increases.

Figure 5(a) shows the results of regret for 8, 64, and 512 slot machines. Regret converges to a larger value as the number of the slot machines increases. Figure 5(b) shows the scalability of the final value of the regret as the number of slot machines is changed. The exponent of 0.923 is obtained as the number of slot machine increases. These values of the power are smaller than those obtained in the literature [3,4].

## 6. Conclusion

In this study, we experimentally generated chaotic spatiotemporal dynamics in an optoelectronic feedback system with a semiconductor laser, a camera, and a spatial light modulator. Chaotic spatiotemporal dynamics were observed by using a large feedback strength β. The generated chaotic spatiotemporal dynamics were applied to a decision-making experiment with a large number of slot machines. We experimentally succeeded in solving the multi-armed bandit problem with up to 512 slot machines. The scalability of the number of plays for CDR = 0.95 and regret were the exponents of 0.902 and 0.923 as the number of slot machine $N$ is changed. These exponents are smaller than those obtained from the previous studies.

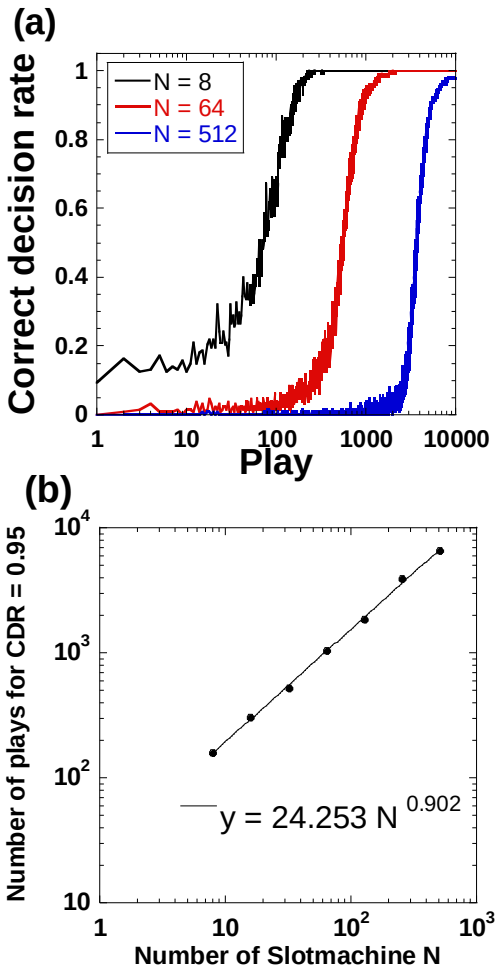


Fig. 4 (a) Results of correct decision rate (CDR) for 8, 64, and 512 slot machines. (b) Scalability of the number of plays for $CDR = 0.95$ as the number of slot machines $N$ is changed.

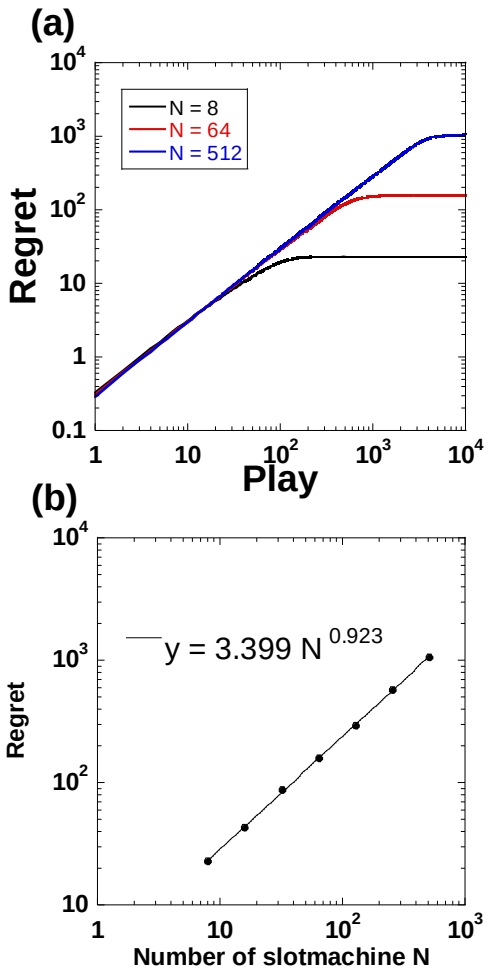


Fig. 5 (a) Results of regret for 8, 64, and 512 slot machines. (b) Scalability of the number of plays for $CDR = 0.95$ as the number of slot machines $N$ is changed.

**Acknowledgement**

This study was supported in part by JSPS KAKENHI (JP19H00868, JP20K15185, JP20H00233, JP22H05195), JST CREST (JPMJCR17N2), and the Telecommunications Advancement Foundation.

**References**

[1] H. Robbins, "Some aspects of the sequential design of experiments," Bulletin of the American Mathmatical Society, Vol. 58, No. 5, pp. 527-536 (1952).

[2] M. Naruse, Y. Terashima, A. Uchida, and S. J. Kim, "Ultrafast photonic reinforcement learning based on laser chaos," Scientific Reports, Vol. 7, Article No. 8772, pp. 1-10 (2017).

[3] M. Naruse, T. Mihana, H. Hori, H. Saigo, K. Okaura, M. Hasegawa, and A. Uchida, "Scalable photonic reinforcement learning by time-division multiplexing of laser chaos," Scientific Reports, Vol. 8, Article No. 10890, pp. 1-16 (2018).

[4] K. Morijiri, T. Mihana, K. Kanno, M. Naruse, and A. Uchida, "Decision making for large-scale multi-armed bandit problems using bias control of chaotic temporal waveforms in semiconductor lasers," Scientific Reports, Vol. 12, Article No. 8073, pp. 1-11 (2022).

[5] S. J. Kim, M. Aono, and E. Nameda, "Efficient decision-making by volume-conserving physical object," New Journal of Physics, Vol. 17, No. 8, pp. 083023 (2015).

[6] S. J. Kim and M. Aono, "Amoeba-inspired algorithm for cognitive medium access," Nonlinear Theory and Its Applications, IEICE, Vol. 5, No. 2, pp. 198-209 (2014).

[7] P. Antonik, N. Marsal, D. Brunner, and D. Rontani, "Human action recognition with a large-scale brain-inspired photonic computer," Nature Machine Intelligence, Vol. 1, pp. 530-537 (2019).

[8] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural network," Optica, Vol. 5, Article No. 6, pp. 756-760 (2018).