HAL
open science

Decision Making Photonics: Solving Bandit Problems Using Photons

Makoto Naruse, Nicolas Chauvet, Atsushi Uchida, Aurelien Drezet, Guillaume Bachelier, Serge Huant, Hirokazu Hori

To cite this version:

Makoto Naruse, Nicolas Chauvet, Atsushi Uchida, Aurelien Drezet, Guillaume Bachelier, et al.. Decision Making Photonics: Solving Bandit Problems Using Photons. IEEE Journal of Selected Topics in Quantum Electronics, 2020, 26 (1), pp.7700210. 10.1109/jstqe.2019.2929217 . hal-03026803

HAL Id: hal-03026803

https://hal.science/hal-03026803

Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Decision Making Photonics: Solving Bandit Problems by Photons

(Invited paper)

Makoto Naruse[1]*, Nicolas Chauvet[1], Atsushi Uchida[2], Aurélien Drezet[3], Guillaume Bachelier[3], Serge Huant[3], and Hirokazu Hori[4]

[1] *Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2] *Department of Information and Computer Sciences, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama, Saitama 338-8570, Japan*

[3] *Univ. Grenoble Alpes, CNRS, Institut Néel, 38000 Grenoble, France*

[4] *Interdisciplinary Graduate School, University of Yamanashi, Takeda, Kofu, Yamanashi 400-8510, Japan*

* Corresponding author. Email: naruse@nict.go.jp

Abstract

Decision making in dynamically changing uncertain environments is one of the most important elements in information and communications technology, ranging from resource assignments in data centers, wireless communications, search functions, among others. Here we review our research on physically efficiently realizing or accelerating decision making by photonics. Specifically, the problem under study is multi-armed bandit problem (MAB) to maximize total rewards from unknown environments that involve difficult trade-off called exploration-exploitation dilemma. We show the principle of solving MAB problems by utilizing the wave-particle duality of single photons in which the probabilistic attribute of single light quanta plays the role of exploration. The principle is transformed to ultrafast laser chaos where the chaotically oscillated irregular time series provides fast and scalable decision-making abilities. The problem becomes even more difficult when multiple players are involved, called competitive MAB (CMAB) problem, in which the expected value could regard to social benefit maximization and ensuring equality or fairness for individuals. We show that entangled photons can resolve CMAB problem. Theoretical studies on photonic decision making are also reviewed showing where total six entities are interacted with each other in a form called octahedron structure.

Index-terms: Decision making, photonics, reinforcement learning, artificial intelligence, multi-armed bandit problem, single photon, laser chaos, entangled photon, categorical system model

I. INTRODUCTION

Decision making is to conduct adequate judgements in dynamically changing uncertain environment, which widely applies in information and communications technology ranging from frequency and channel assignments in networks [Lai, Kuroda], Monte-Carlo tree search in artificial intelligence [Silver], among others. Decision making is also the foundation of reinforcement learning [Sutton]. One of the most important and fundamental issues in decision making is the multi-armed bandit (MAB) problem where the point is to find the best slot machine dispensing a lot of reward from many slot machines whose reward probabilities are unknown [Sutton]. To find the best machine, one must sufficiently explore the machines; however, too much exploration may involve much loss. On the other hand, too quick a decision may miss the best machine. Moreover, the best machine may change overtime known as uncertainty in environments; one must change his/her decision depending on situations. Thus, there is a difficult trade-off called exploration-exploitation dilemma. In this study, decision making refers to solve MAB problems.

We examine solving MAB problems physically with photons and photonic technologies instead of conventional computer algorithms performed in digital computers [Daw, Robbins, Auer] with a view to benefiting from unique physical attributes of photons to pave the way for breaking the limitations of conventional approaches such as von Neumann bottleneck [Backus], energy efficiency, operating speeds, and creating novel values. In particular, by pursuing the ultimate performances of photons, we can design novel system architectures and functionalities. In Sec. II, single-photon

decision maker is shown by utilizing wave-particle duality of light quanta. The principle is, in Sec. III, transformed to ultrafast chaotic lasers where the high bandwidth nature of light is utilized, allowing additionally a superior scalability by time-domain multiplexing of chaotic time series. The MAB problems become even more difficult when multiple decision makers are involved known as competitive MAB (CMAB) problem. The decision conflict inhibits maximum utilizations of available opportunities; namely, social value maximization is expected. Additionally, the issue of equality or fairness of individual players becomes a concern. In Sec. IV, we show the proof-of-principle experimental demonstration of solving CMAB by entangled photons. In photon-based decision makers, various elements ranging from decisions, betting results, probabilistic attributes of photons, etc. are involved; understanding of the complex mechanisms is important for systematic system design. In Sec. V, a theoretical study based on category theory is presented where the notion of triangulated category clearly describes the inherent interdependencies of the subject matter. Section VI concludes the paper.

Research of photonics for intelligent functionalities has been intensively studied from late 2010s [Inagaki, Larger, Brunner, Chen, Bueno] in accordance with a variety of social and technological developments such as ever-growing importance of computing power and artificial intelligence, expected end of Moore's low [Theis], and extensive progress of photonic technologies since the extensive optical computing and optical neural network research in 1980s [Psaltis, Ishikawa]. Photonic decision making shows a clear departure from recent photonic computing such

as photonic Ising machines [Inagaki], optical reservoir computing [Larger, Brunner], and light-based neural networks [Shen, Bueno] in terms of its intended functionalities and architectures technologies. The goal of the other recent systems is combinatorial optimization or recognition/classification tasks involving certain amount of training data sets, which does not apply for photonic decision maker examined in the present study. However, it should be emphasized that the photonic decision makers and photonic solution searchers are in a *complemental* relation. Indeed, for example, Kanno *et al.* demonstrated the composite system of laser-chaos decision maker and optical-fiber-based reservoir computing where dynamic model selection has been demonstrated that enhances the prediction ability of the reservoir computing as a whole [Kanno]. That is, the fusion of various photonic systems, as well as non-photonic systems, with photonic decision making is an interesting future possibility. Such future studies will be briefly discussed at the end of the paper.

II. SINGLE PHOTON DECISION MAKER

The difficulty of MAB becomes evidently clear when the number of candidates increases. However, only two choices, that is, two-armed bandit problem, involves significant difficulties. Intuitively, one may think that it would be reasonable to keep choosing the slot machine when the selected machine mostly wins in recent trials. However, in reality, one can be easily fooled by such incidental events: the unselected, other machine may be the truly higher probability slot machine. That is, maintaining certain moment of *thinking-over* is critical for correct decision making.

The idea of single-photon decision maker is based on the fact that the wave-particle duality of light quanta directly applies for the structure of solving two-armed bandit problem [MN15]. As shown in Fig. 1(a), we utilized single photons emitted from a NV-center in a nondiamond. A single photon impinges on a polarization beam splitter (PBS) after passing through a polarizer. If the photon is detected by APD_A, which is the vertical direction of PBS, the decision is immediately to choose the slot machine A whereas when the photon is measured by APD_B for the horizontal polarization, the decision is to select the slot machine B. A linearly-polarization single photon polarized 45 degrees with respect to the horizontal is detected either by APD_A or APD_B with 50:50 probability. On the other hand, when the polarization of input single photon is almost vertical, the photon is highly likely observed by APD_A whereas almost vertically-polarized single photon will be detected by APD_B. Hence, the strategy of decision making is to control the single photon polarization toward the better slot machine.

What is important is the above-mentioned thinking-over process; this is physically supported by the probabilistic attribute of single photons. For example, a nearly horizontal single photon is mostly detected by APD_B, but sometimes, it is detected by APD_A. We should emphasize that such a property cannot be achieved if the input photon is classical; when the physical quantity measured by photodetectors are the ratio of light intensity, we need one additional step to determine the decision [SR15].

In the experiment, a NV-center in a nanodiamond was excited a green laser. The arrival timing of photons at APDs was characterized by a time-correlated single-photon detection system. The two slot machines were implemented in a computer using pseudorandom numbers. Based on the betting results, the angle of the half-wave plate was mechanically controlled to configure the polarization of single photons. A representative experimental result is shown in Fig 1(b). The horizontal axis shows the time or the number of cycles of slot machine play while the vertical axis depicts correct decision ratio (CDR) which is the ratio of selecting higher reward probability slot machine over total ten repetitions. In the first 150 cycles, the reward probabilities of the slot machine A and B were given by 0.8 (P_A = 0.8) and 0.2 (P_B = 0.2), respectively; hence the selection of slot machine A is *correct* decision. The red solid curve is rapidly approaching to one, meaning that correct decision-making is being conducted. In every 150 cycles, the reward probabilities of the slot machines are swapped in order to implement environmental uncertainty. Accordingly, CDR drops right after the 150 cycle, but it gradually recovers to high scores, indicating that the autonomous adaptation to environmental changes were demonstrated. The blue dotted line shows the results when the reward probabilities are given by 0.6 and 0.4, which is more difficult decision problem since the difference between the reward probability is smaller than the previous case. Accordingly, the CDR exhibits relatively lower scores compared with the former case. Nevertheless, the autonomous decision making was clearly demonstrated.

Before finishing Sec. II, we describe several remarks. First, essentially the same architecture with the above-described single-photon decision maker has been implemented using optical near-field coupling between quantum dots, which is theoretically examined in [SR13] followed by experimental demonstration [JA], pursuing the ultra-small nature of optical near-field in the subwavelength scale [RoPP]. The energy transfer via optical near-field was modulated by external control light so that the photon was transferred preferably to either of the quantum dots corresponding to decisions to select machines A or B. Meanwhile, recent advancements in nano-optics allow further functionalities including memorization of past events, which are important elements in decision making. Nakagomi *et al.* demonstrated in recording nano-scale patterns on the surface of photochromic single crystal via optical near-field [Nakagomi].

Secondly, by cascading PBSs in a tree-form architecture, the four-armed bandit problem has been experimentally resolved using single photons [ACS]. Such scalability of the photon-based decision making is discussed in detail later in Sec. III. Here we put a remark on the control mechanism of the polarizer. Suppose that $P_A = 0.8$ and $P_B = 0.7$, meaning that both slot machine yield rewards rather frequently. No matter what the decision is, one can *win* the slot machine playing by 75% whereas the *lose* events are rare with only 25%. This means that the event of *lose* must be highly, more specifically, three-times largely evaluated than the event of *win*. On the other hands, when $P_A = 0.3$ and $P_B = 0.2$, meaning that both slot machines yield very smaller rewards, one mostly loses by a factor of 75 % with only 25% rate of winning; hence the event of *win* must be three-times

largely appreciated than the event of lose. In such manners, the degree of polarization control is attenuated or accelerated by considering the ratio of the overall winning figure $(P_A + P_B)$ to the overall losing metric $[2-(P_A + P_B)]$ [ACS]. Generally, this implies that the information of environments seriously affects the decision, which is theoretically analysed in the categorical investigation discussed in Sec. IV. Naruse *et al.* also present local reservoir model to account of decision making featuring the effects of environments [PLOS].

III. ULTRAFAST DECISION MAKING BY LASER CHAOS

The single photon and near-field photon decision making discussed in Sec. II directly utilize the single light quanta, demonstrating the ultimate abilities of photon at its energy efficiency and spatial density. However, at the point of our research, there are several technological difficulties such as the operating speed limits of single photon sources or mechanical control of polarizations. On the other hands, one of the significant attributes of light is its ultrahigh bandwidth, which has been widely utilized in long- and short-distance optical communications, for example. This section transforms the principle shown in Sec. II in the time-domain property of lightwave by using ultrahigh speed chaotic lasers.

A. Laser Chaos

The generation of chaos in semiconductor lasers has been extensively studied in the literature [Ohtsubo, UchidaWiley]. For example, the oscillation of lasers becomes unstable, leading to chaos, when a portion of output light is fed back to the laser cavity after certain delay via an externally

arranged mirror. Therefore, antireflection coatings or optical isolators are indispensable to guarantee stable operations of lasers in general applications. However, such laser chaos also provides unique sources of randomness. Indeed, ultrahigh speed physical random generations with a view to accomplish fast rates unachievable by conventional algorithms on digital computers have been experimentally demonstrated [Uchida08, Argyris]. Furthermore, the nonlinear dynamics of chaotic lasers, including its ultrafast transient process, is a fascinating resource for reservoir computing [Brunner]. The application of laser chaos to decision making described herein is one of the most recently proposed and demonstrated utilizations of nonlinear dynamics of lasers for intelligent functionalities.

B. Principle

Abstractly, the principle of the laser-chaos decision maker is equivalent to the single-photon decision maker in Sec. II, but its realization and enabling features completely differ from each other. As shown schematically in Fig. 2(a), a light intensity level is sampled from the laser chaos signal, which is then subjected to a comparison to a given threshold (*TH*) value. When the signal level is larger than *TH*, the decision is immediately determined to select slot machine A whereas the case with signal level being lower than *TH* the decision is to choose slot machine B. The threshold level is reconfigured based on the betting result. Let, for instance, the threshold being sufficiently high. Consequently, the sampled signals result mostly being smaller than the threshold, leading to the decision to take slot machine B. However, thanks to chaotically oscillating input light, the sample

could sometime be *even larger* than the threshold; that is, the decision to select machine A is occasionally taken. In such a manner, the action of thinking-over is physically accomplished as in the case of the single-photon decision maker.

In [SR17], with using a chaotic time series generated by a semiconductor laser with a delayed feedback, ultrafast and adaptive decision-making was demonstrated for solving two-armed bandit problems. A semiconductor laser operated at a center wavelength of 1547.785 nm is coupled with a polarization-maintaining (PM) coupler. The light is connected to a variable fiber reflector which provides delayed optical feedback to the laser. The output light at the other end of the PM coupler is detected by an AC-coupled photodetector through an optical isolator and optical attenuator. The signal is sampled by a high-speed digital oscilloscope at a rate of 100 GSample/s (a 10 ps sampling interval) with an eight-bit resolution.

The blue curve in Fig. 2(b) shows the evolution of CDR when the chaotic signal is sampled with 50 ps interval, namely, at a rate of 20 GSample/s that exhibits the promptest adaptation to the unknown environments. After about 20 cycles of trials, the CDR reaches above 0.9, indicating that the latency from the initial no knowledge status to the correct decision is about 1 ns. The sampling interval 50 ps that provides the best decision-making performance exactly coincides with the negative autocorrelation inherent in the chaotic time series shown in the inset of Fig. 2(b). Meanwhile, although a quasiperiodic signal exhibits larger negative maximum in its autocorrelation the decision-making results in poor performances. Moreover, even assuming that pseudorandom

numbers and color noise were available in such a high-speed domain, the laser chaos outperformed these alternatives (Fig. 2(a)); that is, chaotic dynamics yields superior decision-making abilities. Here, the color noise containing negative autocorrelation was calculated based on the Ornstein–Uhlenbeck process using white Gaussian noise and a low-pass filter [Fox] with a cut-off frequency of 10 GHz. Such an aspect regarding the temporal structure of irregular signals and the decision making is further discussed in the next section.

C. Scalable chaos-based decision maker

Scalability is an important aspects of information systems in general. Here, developing principles and technologies toward an N-armed bandit with N being a large natural number is of great interests. Taking advantage of the high-bandwidth attributes of chaotic lasers, we proposed and demonstrated the time-division multiplexing method in the decision-making strategy; specifically, consecutively sampled chaotic signals were used to determine the identity of the slot machine in a binary digit form.

We considered a MAB problem in which a player selects one of N slot machines, where $N = 2^M$ with M being a natural number. The N slot machines are distinguished by the *identity* given by natural numbers ranging from 0 to $N - 1$, which are also represented in an M-bit binary code given by $S_1 S_2 \mathrm{L}\ S_M$ with S_i (i = 1, …, M) being 0 or 1. For example, when N = 8 (or M = 3), the slot machines are numbered by $S_1S_2S_3$ = {000, 001, 010, ..., 111} (Fig. 3(a)). The reward probability of slot machine i is represented by P_i (i = 0, …, $N - 1$), and the problem addressed herein is the selection of the machine with the highest reward probability.

The identity of the slot machine to be chosen is determined bit by bit from the most significant bit (MSB) to the least significant bit in a pipelined manner. For each of the bits, the decision is made based on a comparison between the measured chaotic signal level and the designated threshold value. First, the chaotic signal $s(t_1)$ measured at $t = t_1$ is compared to a threshold value denoted as TH_1 (Fig. 3(b)). The output of the comparison is *immediately* the *decision* of the MSB concerning the slot machine to choose. If $s(t_1)$ is less than or equal to the threshold value TH_1, the decision is that the MSB of the slot machine to be chosen is 0, which we denote as $D_1 = 0$. Otherwise, the MSB is determined to be 1 ($D_1 = 1$). Here we suppose that $s(t_1) < TH_1$; then, the MSB of the slot machine to be selected is 0. Based on the determination of the MSB, the chaotic signal $s(t_2)$ measured at $t = t_2$ is subjected to another threshold value denoted by $TH_{2,0}$. The first number in the suffix, 2, means that this threshold is related to the *second*-most significant bit of the slot machine, while the second number of the suffix, 0, indicates that the previous decision, related to the MSB, was 0 ($D_0 = 0$). If $s(t_2)$ is less than or equal to the threshold value $TH_{2,0}$, the decision is that the second-most significant bit of the select slot machine to be chosen is 0 ($D_2 = 0$). Otherwise, the second-most significant bit is determined to be 1 ($D_2 = 1$).

All of the bits are determined in this manner. In general, there are 2^{k-1} kinds of threshold values related to the k-th bit; hence, there are $2^M - 1 = N - 1$ kinds of threshold values in total. What is important is that the incoming signal sequence is a chaotic time series which enables efficient exploration of the searching space, as discussed later.

Based on the betting results, the threshold values are adjusted so that that the same decision will be highly likely to be selected in the subsequent play. Therefore, for example, if the MSB of the selected machine is 0, TH_1 should be *increased* because doing so increases the likelihood of obtaining the same decision regarding MSB being 0. All of the other threshold values involved in determining the decision are updated in the same manner. The details of the principle are described in [SR18].

Four kinds of chaotic signal trains were generated, referred to as Chaos 1, Chaos 2, Chaos 3, and Chaos 4 by varying the reflection by the variable reflector by letting 210, 120, 80, and 45 μW of optical power be fed back to the laser, respectively. While the time-domain waveforms of Chaos 1–4 look similar, there was a clear difference in their radio-frequency power spectra obtained using Chaos 1, 2, 3, and 4, as shown in [SR18]. A quasiperiodic signal train was also generated by the variable reflector by providing a feedback optical power of 15 μW. In addition, computer-generated, uniform pseudorandom numbers (RAND) and color noise, of which generation is the same the one in Sec. II.B, were examined for comparisons.

We applied the proposed time-division multiplexing decision-making strategy to bandit problems with two, four, eight, 16, 32, and 64 arms. We assigned the reward probabilities to the multiple slot machines so that the difficulty keeps coherence. First, the highest and the second highest reward probabilities were given by given by 0.9 and 0.7, respectively. Second, the probabilities were arranged so that the *contradiction* condition applies for all group of the slot

machines. In the case of four-armed bandit, for example, the probabilities were given by (P_0, P_1, P_2, P_3) = (0.7, 0.9, 0.5, 0.1) where the maximum-reward-probability machine is machine 1. Neglecting the LSB, the machines are grouped to {Machine 0 and Machine 1} and {Machine 2 and Machine 3} where the sum of the probabilities is larger in the latter group although the best machine belongs to the former group, which we call *contradiction* condition. Such situations are satisfied for all groups for the sake of coherent comparison with the increased arm numbers. The details are described in [SR18].

Figures 3(b) summarize the results of the 16-, 32-, and 64-armed bandit problems, respectively. The red, green, blue, and cyan curves show the CDR evolution obtained using Chaos 1, 2, 3, and 4, respectively, while the magenta, black, and yellow curves depict the evolution obtained using quasiperiodic signals, pseudorandom numbers and color noise, respectively. From Fig. 3(b), it can be seen that Chaos 3 provides the promptest adaptation to the unity value of the CDR, whereas the nonchaotic signals (quasiperiodic, RAND, and color noise) yield substantially deteriorated performances. The number of cycles necessary to reach a CDR of 0.95 increases as the number of bandits in the form of the power-law relation aN^b, where a and b are approximately 52 and 1.16, respectively, indicating that the successful operation of the proposed scalable decision-making principle using laser-generated chaotic time series.

In the results shown for bandit problems with up to 64 arms, Chaos 3 provides the best performance among the four kinds of chaotic time series. The negative autocorrelation indeed affects

the decision-making ability, as discussed in Sec. III.B; however, the value of the negative maximum of the autocorrelation does not coincide with the order of performance superiority, indicating the necessity of further insights into the underlying mechanisms. In this respect, we analysed the diffusivity of the temporal sequences based on the ensemble averages of the time-averaged mean square displacements (ETMSDs) [Miyaguchi, SR16]. A random walker is generated via comparison between the chaotic time series and a uniformly distributed random number; when the chaotic signal $s(t)$ is larger than the random number, the walker moves to the right $X(t) = +1$ otherwise, $X(t) = -1$. Hence, the position of the walker at time t is given by $x(t) = X(1) + X(2) + \ldots + X(t)$. We then calculate the ETMSD using

$$ETMSD(\tau) = \left\langle \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(x(t+\tau) - x(t) \right)^2 \right\rangle, \tag{1}$$

where $x(t)$ is the time series, T is the last sample to be evaluated, and $\langle \mathrm{L} \rangle$ denotes the ensemble average over different sequences. The ETMSDs corresponding to Chaos 1, 2, 3, and 4 and quasiperiodic, RAND, and color noise at the time difference of $\tau = 1000$ exhibits the maximum value followed by Chaos 2, 1, and 4 as shown in Fig. 4(a). This order agrees with the superiority order of the decision-making performance in the 64-armed bandit problem shown in Fig. 3(b). At the same time, RAND and colour noise exhibit larger ETMSD values than Chaos 1–4, although the decision-making abilities are considerably poorer for RAND and coloured noise, implying that the ETMSD alone cannot perfectly explain the performances.

Figure 4(b) explains diffusivity in another way, where the average displacement $\langle x(t) \rangle$ and $\langle x(t+D) \rangle$ are plotted for each time series superimposed in the *XY* plane with D = 10,000. Although the quasiperiodic and color noise, shown by the magenta and yellow curves, respectively, move toward positions far from the Cartesian origin, their trajectories are biased toward limited coverage in the plane. Meanwhile, the trajectories of the chaotic time series cover wider areas, as shown by the red, green, blue, and cyan curves. The trajectories generated via RAND, shown by the black curve, remain near the origin.

To quantify such differences, we evaluated the covariance matrix $\Theta = \mathrm{cov}(X_1, X_2)$ by substituting $x(t)$ and $x(t + \mathrm{D})$ for X_1 and X_2, where the *ij*-element of Θ is defined by $\frac{1}{N-1}\sum_{i=1}^{N}\left(X_1 - \overline{X_1}\right)\left(X_2 - \overline{X_2}\right)$, with N denoting the number of samples and $\overline{X_i}$ denoting the average of X_i. The *condition number* of Θ, which is the ratio of the maximum singular value to the minimum singular value, indicates the uniformity of the sample distribution. A larger condition number means that the trajectories are skewed toward a particular orientation, whereas a condition number closer to unity indicates uniformly distributed data. The square marks in Fig. 4(a) show the calculated condition numbers where Chaos 1–4 achieve smaller values whereas the quasiperiodic and coloured noise yield larger scores. Through these analyses using the ETMDSs and condition numbers related to the diffusivity of the time series, a clear correlation between the greater diffusion properties inherent in laser-generated chaotic time series and the superiority in the decision-making ability is observable.

IV. ENTANGLEMENT PHOTONS FOR SOCIAL MAXIMUM AND EQUALITY

The decision-making problems under study in previous sections regard to a single player's reward maximization. The problem becomes even more difficult when the number of individuals who join the game is multiple because interest conflicts may easily be induced which deteriorates the amount of each player's reward if the total amount of dispensed reward remains the same, as schematically shown in Fig. 5(a). If the players try to maximize the total reward as a *team*, conflicts of decisions must be avoided to benefit from the potentially achievable total benefits. The problem is referred to as a *competitive multi-armed bandit* (CMAB) problem, which underlies important practical issues ranging from traffic jam in roads to congestions in information networks [Kai, Kim16, Liu].

In this section, we demonstrate the proof-of-principle experimental demonstrations of showing the usefulness and superiority of entangled photons for collective decision making [arXiv]. For the simplest case that preserves the essence of the CMAB problem, we consider two players (called Players 1 and 2), each of whom selected one of two slot machines (Machines A and B), with the goal of maximizing the total team reward. The amount of reward that could be dispensed by each slot machine per play is assumed to be unity; hence, when the two players make the same decision, the reward is divided into two halves. This example manifests that players could be easily locked in a local minimum due to conflicts between their decisions since everyone wants more rewards and tries

to select the higher-reward-probability slot machine whereas the total team rewards could be increased if they cooperated.

We utilized entangled photons for the collective decision making. The signal and idler photon of an entangled photon pair regards to the decision of Players 1 and 2, respectively, as schematically shown in Fig. 5(a). Experimentally, the entangled photons were generated based on a standard Sagnac loop architecture [Fedrizzi] used to generate the photon states by spontaneous parametric down conversion. In the branch corresponding to Player 1, each signal photon is subjected to a PBS (PBS_1); if the photon is detected by the avalanche photodiode corresponding to the horizontally polarized light ($APD1_H$), also called H-photon detection hereafter, the decision of Player 1 is to choose Machine A, whereas if the photon is detected by $APD1_V$ corresponding to the vertical polarization (V-photon), then the decision of Player 1 is to choose Machine B. The same hold for player B by exchanging 1 by 3 and 2 by 4. Note that the two slot machines are externally arranged: we emulate the slot machines in a computer using pseudorandom sequences.

The significance of entangled photons is that when Player 1 detects H-photon, Player 2 detects V-photon whereas when Player 1 detects V-photon, Player 2 detects H-photon when the photon pair is represented by $\frac{1}{2}\left(|HV\rangle - |VH\rangle\right)$ state known as the maximally entangled singlet photon state. That is, no decision conflicts result.

In the experiments, the reward probabilities of Machines A and B are given by $P_A = 0.2$ and $P_B = 0.8$, respectively, for the first 50 plays. In the next 50 plays, the reward probabilities are

swapped, i.e. P_A = 0.8 and P_B = 0.2, to emulate a variable environment. Therefore, from the standpoint of *individual* players, selecting Machine B is the *correct* decision in the first 50 plays and taking Machine A for the last 50 plays is good decision since it is highly likely to gain a greater reward.

When the strategy of the single-photon decision maker demonstrated in Sec. II is implemented for both Players 1 and 2 by updating the halfwave plate located in front of the PBSs (not shown in Fig. 5(a)), the CDR of both Players 1 and 2, shown by the red and blue curves in Fig. 5(b), quickly approaches unity, meaning that both players do choose the higher-reward-probability machine in the first half. At cycle 51, the CDR drops due to the flip of the reward probabilities; however, the CDR gradually returns to unity as time elapses, which clearly indicates that both players detect the environmental change and revises their decisions to the higher-reward-probability machine.

However, this result means that both players make the same decision; indeed, the *conflict ratio*, defined as the number of times that the decisions of Players 1 and 2 are identical over the 10 repetitions, exhibits high values close to unity, as shown by the red curve in Fig. 5(c). Consequently, the accumulated rewards of Players 1 and 2 shown by the red and blue curves, respectively, in Fig. 5(d) are seriously decreased compared with the case when only single player play the slot machine (not shown in Fig. 5). The summation of the accumulated rewards of Players 1 and 2, referred to as the *team* reward, is depicted by the green curve in Fig. 5(d), is 70.9 at cycle 100. With the use of

entangled photons, on the other hands, the individual rewards and the team reward are enhanced since the decision conflicts are avoided as demonstrated by the red curve in Fig. 5(c) exhibiting low figures. The team reward reaches at 93.4 at cycle 100, which is almost the theoretical maximum, demonstrating that the entangled photons work for social maximum benefits.

Entangled photons also provide interesting attributes from the viewpoint of equality, which shows contrasting behavior to the case of *correlated* photon pair. Let photon pairs subjected to PBS_1 and PBS_2 are *orthogonally* polarized. With such *correlated* photons, decision conflicts are avoided if the input light for Player 1 is H-photon while at the same time the input for Player 2 being V photon, or vice versa. However, the players always select the same machine, meaning that a specific player always takes the better machine. That is, *equality* or *fairness* is severely deteriorated. Additionally, when the input photons are not exactly configured horizontally or vertically (for example 45 degree), decision conflicts do occur by orthogonally polarized photon pairs.

Conversely with entangled photons, maximized team rewards and equality are *always* guaranteed regardless of the common polarization basis. This is due to the maximally entangled state that is invariant upon rotation of the basis, provided that the bases are the same for both players. Indeed, the CDRs of Players 1 and 2 always randomly fluctuate around 0.5 as shown in Fig. 5(b). This fluctuation agrees with the fact that nearly identical rewards were received by Players 1 and 2, as also observed in Fig. 5(d).

An important condition for establishing the social maximum by polarization-entangled photons is to share the polarization basis among the players. In other words, when the polarization bases are misaligned, team rewards are deteriorated. One can also interpret such a property that the action of *deception* or *outperforming* the other player with the intention of gaining a greater reward *cannot* be accomplished. The equality is preserved with the misaligned polarization bases with deteriorated social rewards. On the contrary, deception or greedy action by a player is achievable when the system is governed by correlated photons. The detailed theory and experimental demonstrations are shown in [arXiv].

V. CATEGORY THEORETIC ANALYSIS OF PHOTON DECISION MAKING

Photon-based decision-making systems are operated with a variety of elements interacted with complex interdependencies. The elements include dynamically changing uncertain environments, namely, probabilistic attributes of slot machines and changes of probabilities, irregular nature of lights such as the probabilistic attributes of single photons, complex waveforms of laser chaos, and the controllers such as polarization controller or threshold adjusting mechanisms. In view of advanced system developments, systematic ways of system design, applications, constructing theoretical foundations and clearly grasping the entire system structure is indispensable. We should also emphasize that a part of system's functionality is clearly outsourced to physically uncertain

processes, such as in single photons or chaotic lasers. Novel design frameworks and principles are expected. In this section, we review the outline of our approaches toward physical decision making by category theory. See [IJDM] for details. For category theory, well-known textbooks would include [MacLane, Awodey, Simmons].

Here we examine the single-photon decision maker dealing with two-armed bandit problem. The reward probability setting is modeled as an object called "Casino Setting" denoted by X while the setting of the halfwave plate is denoted by an object referred to as "Polarizer Setting" denoted by Y. In [IJDM], it has been demonstrated that, by a simple geometrical calculation, the polarization update strategy discussed in Sec. II autonomously configures Y so that it is approaching to X. That is, correct decision-making results. However, such an analysis does *not* explicitly deal with the randomness inherent in photons and environments; more general treatment is necessary. Categorical approach provides us to deeper understanding of the inherent structure. Let denote each decision (to choose Machine A or B) by an object P while the betting result (win or lose) by an object Q. Intuitively, the decision P directly leads to the result Q; which is represented by $P \to Q$. However, such interpretation prevents us from structural understandings. First, we need to consider that the Casino Setting X is a direct product of P and Q; that is, $X = P \times Q$. Intuitively speaking, since the broker of the casino knows the exact reward probability of the slot machines, he/she can predict the betting result of the player's decision. On the other hand, the Polarizer Setting Y is a direct sum of P

and Q; that is, $Y = P + Q$, indicating that the player can revise his/her decision based on the betting results.

Furthermore, two additional objects are derived. One is the source of probabilistic attributes of the slot machine, referred to as "Machine Environment" denoted by M. It should be emphasized that M *cannot* be observed from the betting result Q. Namely, M is the kernel of Q. Using the notion of homological algebra, such a relation is represented by a short exact sequence given by $0 \to M \to X \to Q \to 0$ [IJDM]. One more object is the source of probabilistic attributes of single photons, named as "Photon Environment" denoted by F. Remember that the photon detection probability is biased by the orientation of the halfwave plate Y. However, if the movement of the halfwave plate is *blocked* or *limited* to some degrees, decision making cannot be improved based on the betting results (Q). In other words, *the room for growth* is important, which mathematically corresponds to the co-kernel of Q; $0 \to Q \to Y \to F \to 0$ [IJDM]. Furthermore, based on the notion of triangulated category [Iversen], a total of six objects (X, Y, P, Q, M, F) are related with each other in the form of octahedron structure which is shown in Fig. 6(a). Here, the wiggled arrows are called characteristic arrow, meaning that the object of the origin affect the next step of the destination object. There are four updating triangles are observed in the octahedron: (Braid 1) $M \to Q \to Y \to M[1]$, (Braid 2) $X \to Q \to F \to X[1]$, (Braid 3) $P \to Y \to F \to P[1]$, (Braid 4) $M \to X \to P \to M[1]$, where $A[1]$ means that the time step of object A is proceeded by a single step. Consequently, by temporally describing the relations in the horizontal direction, the four braids are

evolving through interacting with each other as shown in Fig. 6(b), called braid structure [Iversen]. The physical meaning of the intersections of the braids were examined in [IJDM].

Such a categoric analysis has also been adapted to solution searching system using optical energy transfer between quantum dots involving spatiotemporal probabilistic behavior [Phil]. The octahedron structure applies in the same way as the decision-making instance. A critically important condition for the validity of octahedron structure is that the objects are related in the form of *short exact sequence*, such as $0 \to M \to X \to Q \to 0$ in the case of single-photon decision maker. Physically, this implies that the system should evolve to the next step after realizing an equilibrium state. Conversely, for example, if the system under study are operated too fast, the short exact sequence does not hold; hence, correct solutions cannot be found by the system [Phil]. We proposed a novel notion of time, what we call "*short-exact-sequence-based-time*", to characterize the flow of time for given functionality [Phil]. For example, the solution searching performance in quantum-dot-based device was analyzed as a function of operating speed by which we can quantitatively derive the short-exact-sequence-based time.

The categorical approach is more vividly applied to composite systems that involve physical substrates and environments. Saigo *et al.* utilized *natural transformation*, known to be one of the core concepts in category theory [MacLane], by which rigorously describes the attributes of *soft robots* with mathematically evident difference to conventional hard robots [Saigo]; the softness and its function is characterized by categorical equivalence where abundant degrees of freedom inherent

in soft materials are represented. The probabilistic attributes of photons utilized for decision making would contain similar (or the same) architecture characterized by natural transformation; this is one of interesting future studies.

VI. CONCLUSION

We reviewed our theoretical and experimental studies on photonic decision making. By utilizing the intrinsic physical attributes of single photons, near-field optical interactions, chaotic lasers, multi-armed bandit problems were resolved highlighting the probabilistic attributes of single photons, ultrahigh spatial density of near-field coupling, and ultrahigh speed chaotic dynamics of lasers, respectively. Entangled-photon-based collective decision making is also shown to maximize total team reward as well as ensuring equality among players. In order to systematic understanding of underlying mechanisms of photon-based decision making that involves uncertain natural processes as well as uncertain external environments, category theoretic approach was reviewed where the notion of triangulated category reveals the complex interdependencies of the subject matter.

The research of decision making by photonics is an emergent field; there are many important issues and associated topics to be examined in future. The experimental demonstrations shown in this paper do *not*, of course, cover the potential abilities of photons and photonic technologies for decision making. Simply by limiting the discussion only with chaotic lasers in Sec. III, nonlinear dynamics of lasers allows versatile phenomena, such as synchronization, not just the irregular waveform generated by a simple delayed feedback configuration. The recent growth of photonic

integrated circuits impacts the design of system and device architectures. Deeper considerations into theoretical fundamentals and applications of photonic decision making are also exciting and important areas of research.

ACKNOWLEDGMENT

The authors appreciate many collaborators of our research described in the paper. Especially, the authors thank Martin Berthel of Univ. Grenoble Alpes, Kazutaka Kanno and Takatomo Mihana of Saitama University, Hayato Saigo from Nagahama Institute of Bio-Science and Technology. This work was supported in part by Japan Science and Technology Agency through CREST project (JPMJCR17N2), Japan Society for the Promotion of Science through Core-to-Core Program A. Advanced Research Networks and Grants-in-Aid for Scientific Research (A) (JP17H01277), Agence Nationale de la Recherche through TWIN (Grant No. ANR-14-CE26-0001-01-TWIN) and Placore (Grant No. ANR-13-BS10-0007-PlaCoRe) projects, Université Grenoble Alpes, and Laboratoire d'excellence LANEF in Grenoble through ANR-10-LABX-51-01.

REFERENCES

1. L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation, and competition," *IEEE Trans. Mobile Computing*, vol. 10, no. 2, pp. 239–253, 2011.

2. K. Kuroda, H. Kato, S.-J. Kim, M. Naruse, and M. Hasegawa, “Improving throughput using multi-armed bandit algorithm for wireless LANs,” *Nonlinear Theory and Applications, IEICE*, vol. 9, pp. 74-81, 2018.
3. D. Silver, *et al.* “Mastering the game of go without human knowledge,” *Nature*, vol. 550, pp. 354, 2017.
4. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (The MIT Press, Massachusetts, 1998).
5. N. Daw, J. O’Doherty, P. Dayan, B. Seymour, and R. Dolan, “Cortical substrates for exploratory decisions in humans,” *Nature*, vol. 441, pp. 876–879, 2006.
6. H. Robbins, “Some aspects of the sequential design of experiments,” *B. Am. Math. Soc.* Vol. 58, pp. 527–535, 1952.
7. P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multi-armed bandit problem,” *Machine Learning*, vol. 47, pp. 235–256, 2002.
8. J. Backus, “Can programming be liberated from the von Neumann style?: a functional style and its algebra of programs,” *Commun. ACM*, vol. 21, no. 8, pp. 613–641, 1978.
9. T. N. Theis and H. S. P. Wong, “The end of moore's law: A new beginning for information technology,” *Computing in Science & Engineering*, vol. 19, no. 2, pp. 41, 2017.
10. D. Psaltis, D. Brady, and K. Wagner, “Adaptive optical networks using photorefractive crystals,” *Appl. Opt.*, vol. 27, no. 9, pp. 1752–1759, 1988.

11. M. Ishikawa, N. Mukohzaka, H. Toyoda, and Y. Suzuki, “Optical associatron: a simple model for optical associative memory,” *Appl. Opt.*, vol. 28, no. 2, pp. 291–301 (1989).

12. T. Inagaki, *et al.* “A coherent Ising machine for 2000-node optimization problems,” *Science*, vol. 354, no. 6312, pp. 603–606, 2016.

13. L. Larger, *et al.* “Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing,” *Opt. Express*, vol. 20, pp. 3241–3249, 2012.

14. D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, “Parallel photonic information processing at gigabyte per second data rates using transient states,” *Nat. Commun.*, vol. 4, pp. 1364, 2013.

15. Y. Shen, *et al.* “Deep learning with coherent nanophotonic circuits,” *Nat. Photon.*, vol. 11, no. 7 pp. 441, 2007.

16. J. Bueno, *et al.* Reinforcement learning in a large-scale photonic recurrent neural network. *Optica*, vol. 5, pp. 756–760, 2018.

17. K. Kanno, M. Naruese, and A. Uchida, “Optical reservoir computing with combination of reinforcement learning,” In *The 79th Japan Society of Applied Physics Autumn Meeting*, Abstract, p. 03-139, 2018.

18. M. Naruse, *et al.* “Single-photon decision maker,” *Sci. Rep.* vol. 5, pp. 13253, 2015.

19. S. -J. Kim, M. Naruse, M. Aono, M. Ohtsu, and M. Hara, “Decision Maker Based on Nanoscale Photo-Excitation Transfer,” *Sci Rep.*, vol. 3, pp. 2370, 2013.

20. M. Naruse, *et al.* “Decision making based on optical excitation transfer via near-field interactions between quantum dots,” *J. Appl. Phys.*, vol. 116, pp. 154303, 2014.

21. M. Naruse, N. Tate, M. Aono, and M. Ohtsu, “Information physics fundamentals of nanophotonics,” *Rep. Prog. Phys.*, vol. 76, pp. 056401, 2013.

22. R. Nakagomi, K. Uchiyama, H. Suzui, E. Hatano, K. Uchida, M. Naruse, and H. Hori, “Nanometre-scale pattern formation on the surface of a photochromic crystal by optical near-field induced photoisomerization,” *Sci. Rep.*, vol. 8, pp. 14468, Sep. 2018.

23. M. Naruse, M. Berthel, A. Drezet, S. Huant, H. Hori, and S. -J. Kim, “Single Photon in Hierarchical Architecture for Physical Decision Making: Photon Intelligence,” *ACS Photonics*, vol. 3, pp. 2505–2514, 2016.

24. M. Naruse, E. Yamamoto, T. Nakao, T. Akimoto, H. Saigo, K. Okamura, I. Ojima, G. Northoff, and H. Hori, “Why is the environment important for decision making? Local reservoir model for choice-based learning,” *PLoS ONE*, vol. 13, no. 10, pp. e0205161, October 2018.

25. J. Ohtsubo, *Semiconductor lasers: stability, instability and chaos* (Springer, Berlin, 2012).

26. A. Uchida, *Optical communication with chaotic lasers: applications of nonlinear dynamics and synchronization* (Wiley-VCH, Weinheim, 2012).

27. A. Uchida, *et al.* Fast physical random bit generation with chaotic semiconductor lasers. *Nat. Photon.*, vol. 2, pp. 728–732, 2008.

28. A. Argyris, *et al.* “Implementation of 140 Gb/s true random bit generator based on a chaotic photonic integrated circuit,” *Opt. Exp.*, vol. 18, pp. 18763–18768, 2010.

29. M. Naruse, Y. Terashima, A. Uchida, S. -J. Kim, “Ultrafast photonic reinforcement learning based on laser chaos,” *Sci. Rep.*, vol. 7, pp. 8772, 2017.

30. R. F. Fox, I. R. Gatland, R. Roy, G. Vemuri, “Fast, accurate algorithm for numerical simulation of exponentially correlated colored noise,” *Phys. Rev. A*, vol. 38, pp. 5938–5940, 1988.

31. M. Naruse, T. Mihana, H. Hori, H. Saigo, K. Okamura, M. Hasegawa, and A. Uchida, “Scalable photonic reinforcement learning by time-division multiplexing of laser chaos,” *Sci. Rep.*, vol. 8, pp. 10890, July 2018.

32. T. Miyaguchi and T. Akimoto, “Anomalous diffusion in a quenched-trap model on fractal lattices,” *Phys. Rev. E*, vol. 91, pp. 010102, 2015.

33. S. -J. Kim, M. Naruse, M. Aono, H. Hori, and T. Akimoto, “Random walk with chaotically driven bias,” *Sci. Rep.*, vol. 6, pp. 38634, 2016.

34. S. -J. Kim, M. Naruse, and M. Aono, “Harnessing the Computational Power of Fluids for Optimization of Collective Decision Making,” *Philosophies*, vol. 1, pp. 245–260, 2016.

35. K. Liu and Q. Zhao, “Distributed learning in multi-armed bandit with multiple players,” IEEE Trans. Signal Processing, vol. 58, no. 11, pp. 5667-5681, 2010.

36. M. Naruse, N. Chauvet, D. Jegouso, B. Boulanger, H. Saigo, K. Okamura, H. Hori, A. Drezet, S. Huant, G. Bachelier, Entangled photons for competitive multi-armed bandit problem: achievement of maximum social reward, equality, and deception prevention, arXiv:1804.04316

37. A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, and A. Zeilinger, "A wavelength-tunable fiber-coupled source of narrowband entangled photons," *Opt. Express*, vol. 15, pp. 15377–15386, 2007.

38. M. Naruse S. -J. Kim, M. Aono, M. Berthel, A. Drezet, S. Huant, H. Hori, "Category Theoretic Analysis of Photon-based Decision Making," *International Journal of Information Technology & Decision Making*, vol. 17, no. 5, pp. 1305-1333, May 2018.

39. S. Mac Lane, *Categories for the Working Mathematician* (Springer, Berlin, 1971).

40. S. Awodey, *Category Theory* (Oxford University Press, Oxford, 2010).

41. H. Simmons, *An Introduction to Category Theory* (Cambridge University Press, Cambridge, 2011).

42. B. Iversen, *Cohomology of Sheaves* (Springer-Verlag, Berlin, 1986).

43. M. Naruse, M. Aono, S.-J. Kim, H. Saigo, I. Ojima, K. Okamura, and H. Hori, "Category Theory Approach to Solution Searching based on Photoexcitation Transfer Dynamics," *Philosophies*, vol. 2, no. 3, pp. 16, July 2017.

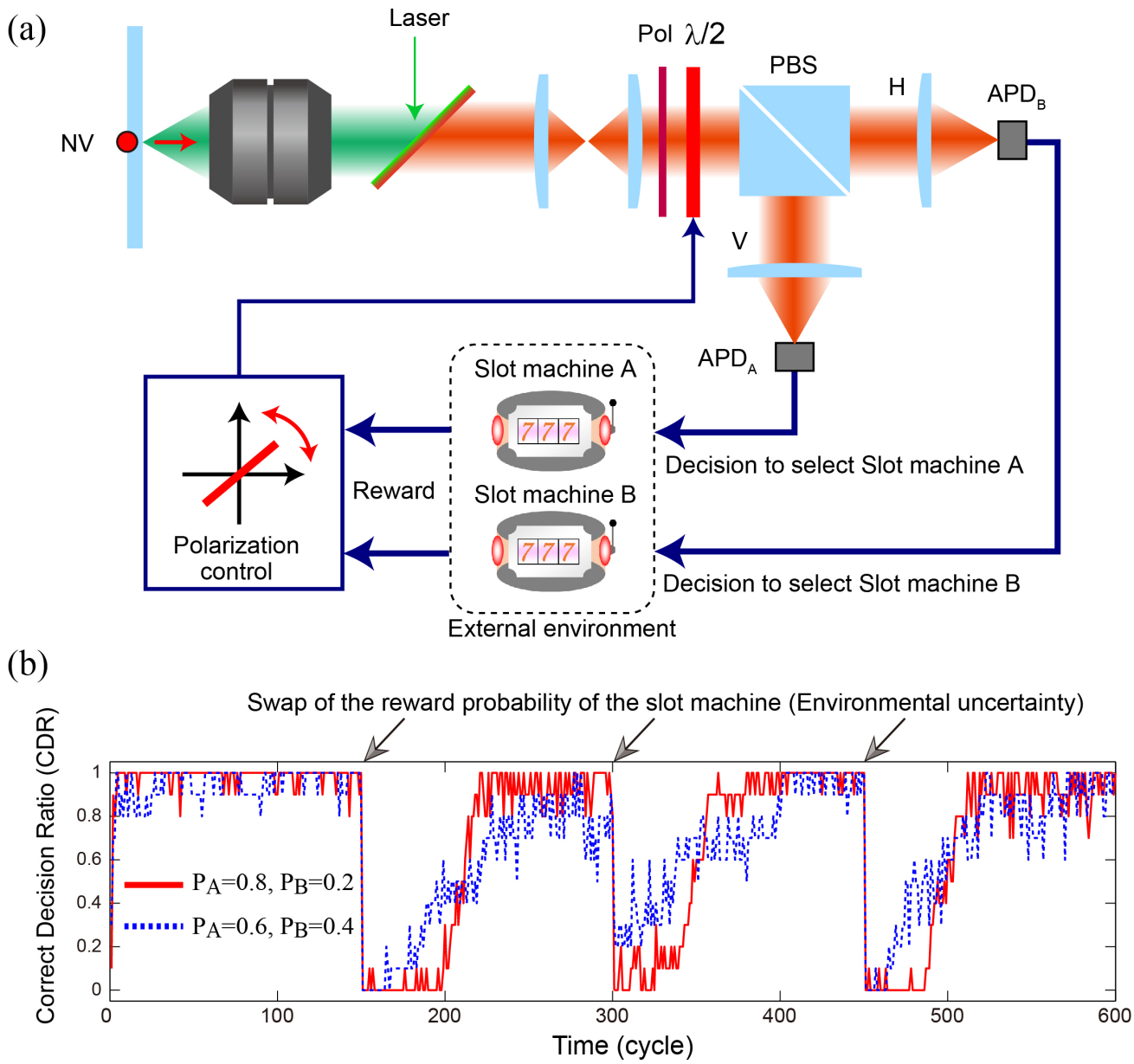

Fig. 1. Single-photon-based decision maker. (a) Architecture of single photon decision maker to solve two-armed bandit problem where the polarization of single photon is controlled. (b) Experimental demonstrations of autonomous decision-making adapting to dynamically changing environment.

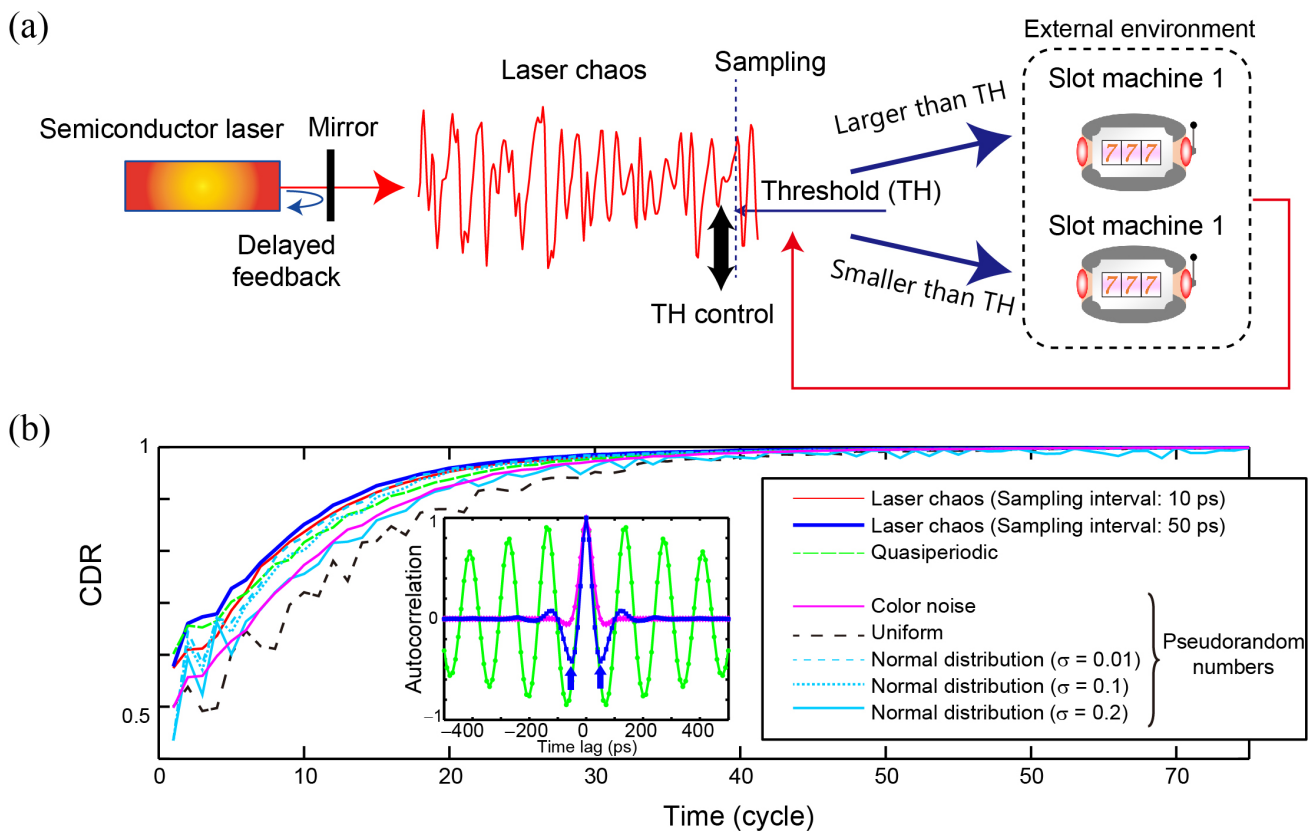

Fig. 2. Laser-chaos-based decision maker. (a) Thresholding to the irregular waveform generated by chaotic lasers directly provides the decision. (b) Experimental demonstrations of solving two-armed bandit problem from zero prior knowledge. Laser chaos sampled with 50 ps interval yields the fastest adaptation, while the autocorrelation of the chaotic time series exhibits negative maximum at the time lag of 50 ps.

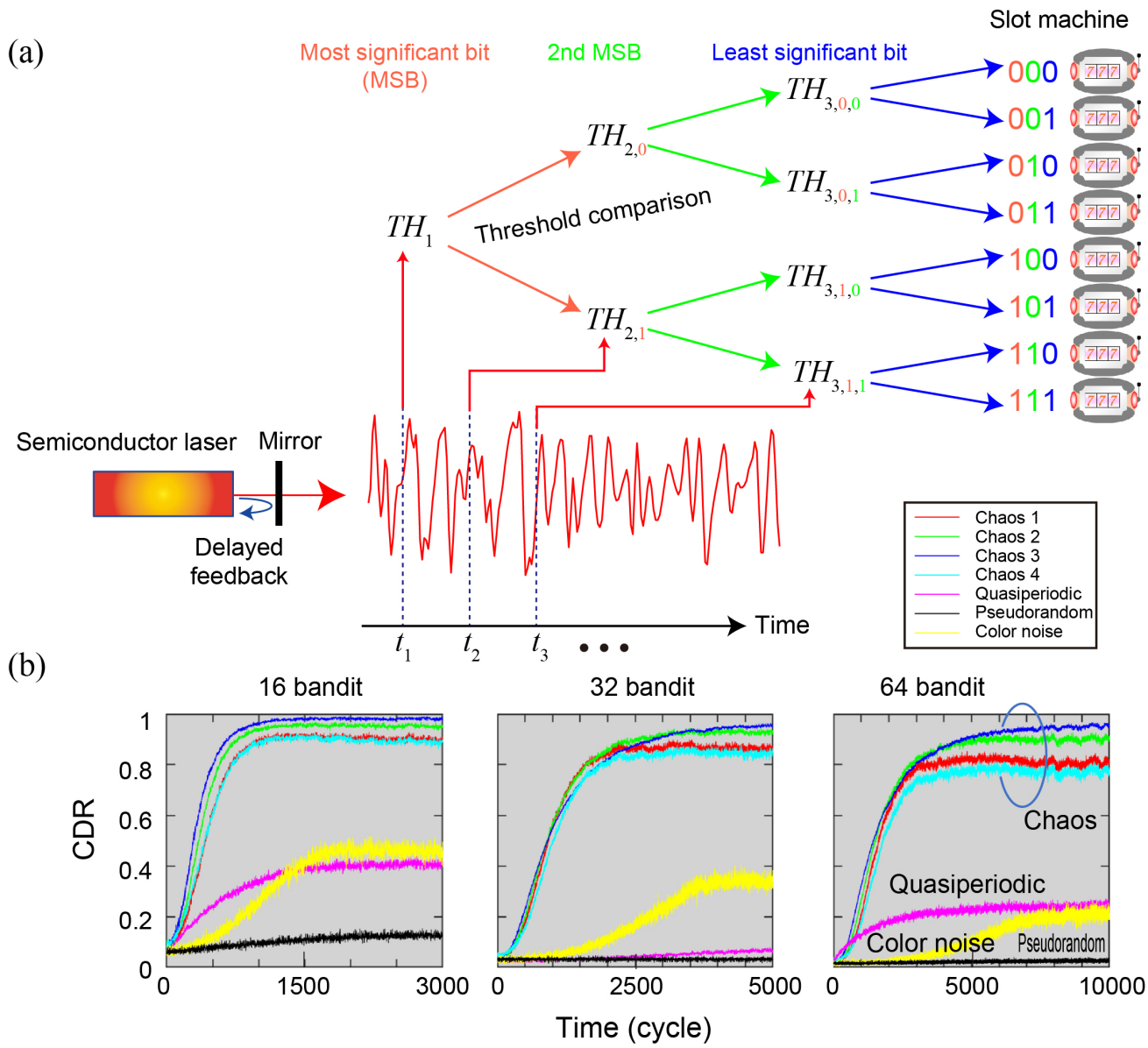

Fig. 3. Scalable decision making by time-domain multiplexing of laser chaos. (a) Schematic diagram of the principle. (b) Solving 16-, 32, and 64-armed bandit problem by chaotic time series and other random signals.

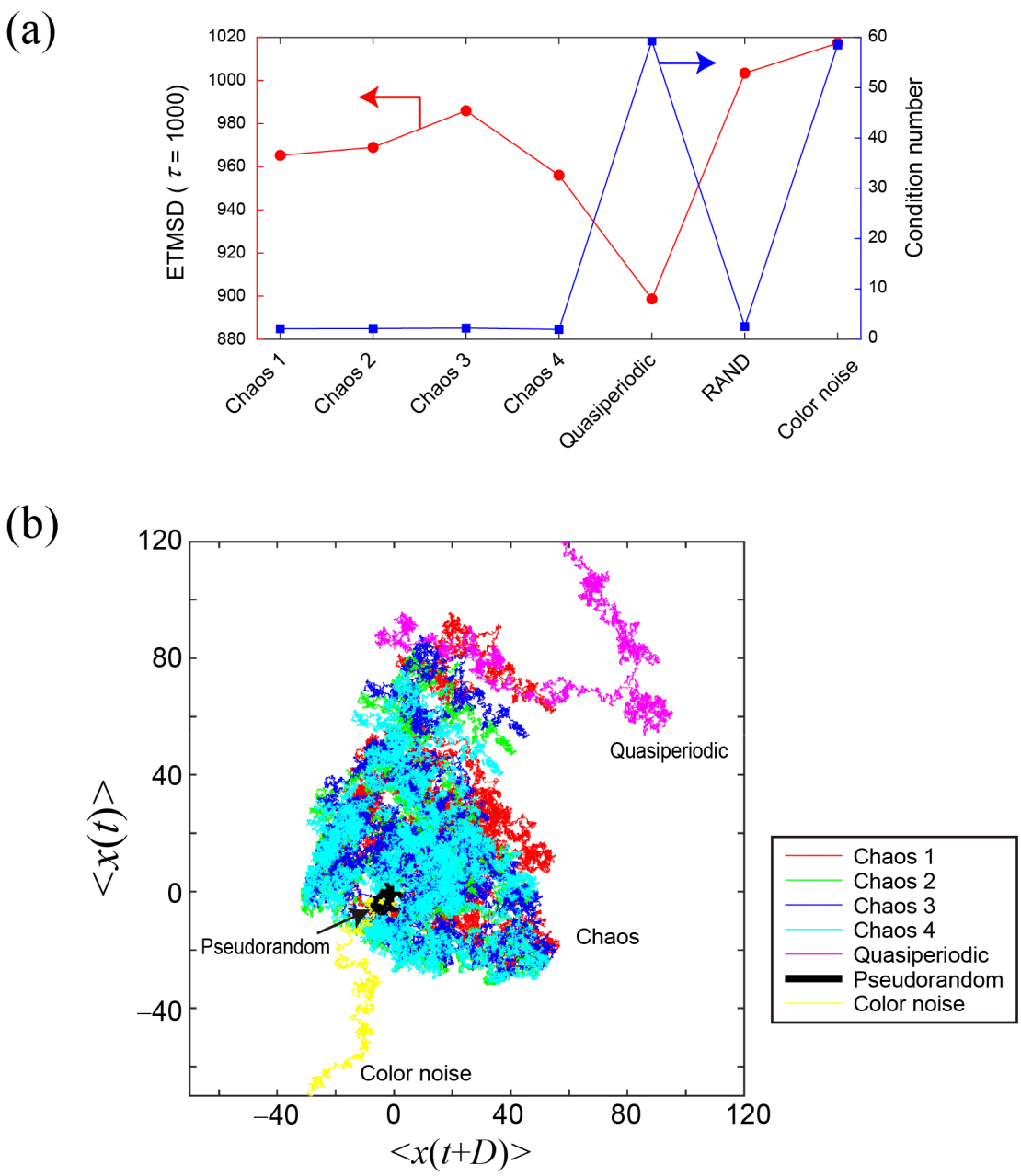

Fig. 4. Diffusivity analysis of chaotic time series. (a) The order of the ensemble-average of mean square displacement agrees with the performance of the decision making among Chaos 1, 2, 3, and 4. (b) A phase map of random walker generated by referring to the given time series. The walker by pseudorandom numbers stays at the origin whereas the ones by quasiperiodic and color noise go far away from the origin but following the same trajectory. The walker by chaotic lasers are distributed in a wider area.

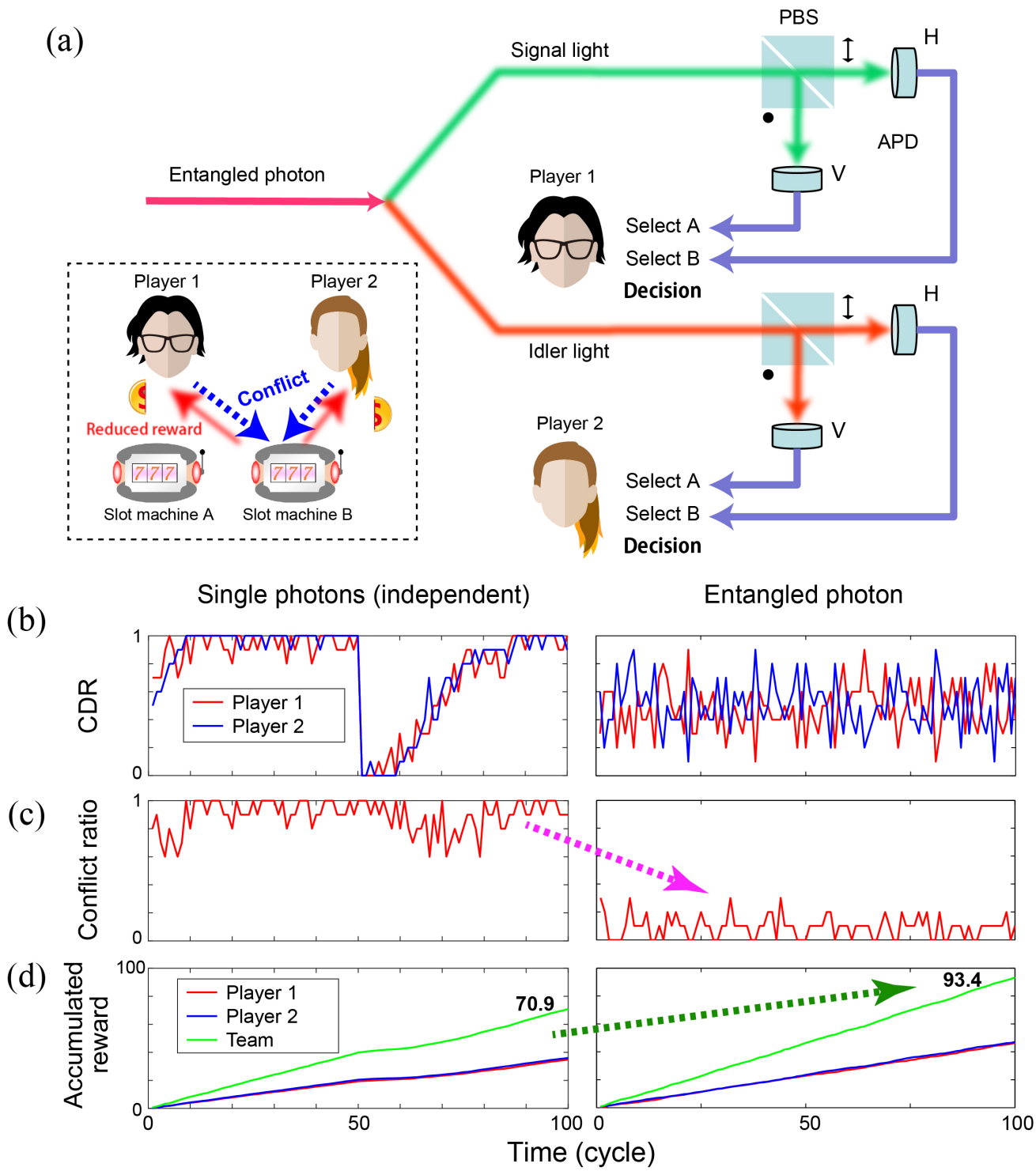

Fig. 5. Collective decision making by entangled photons. (a) Architectural illustration of solving competitive bandit problem by entangled photons. (b, c, d) Experimental demonstration of decision making by two players based on of two independent single photons and entangled photons. (b)

Comparison of collect decision ratio, (c) Decision conflict ratio, and (d) Accumulated total reward by individual players and team as a whole.

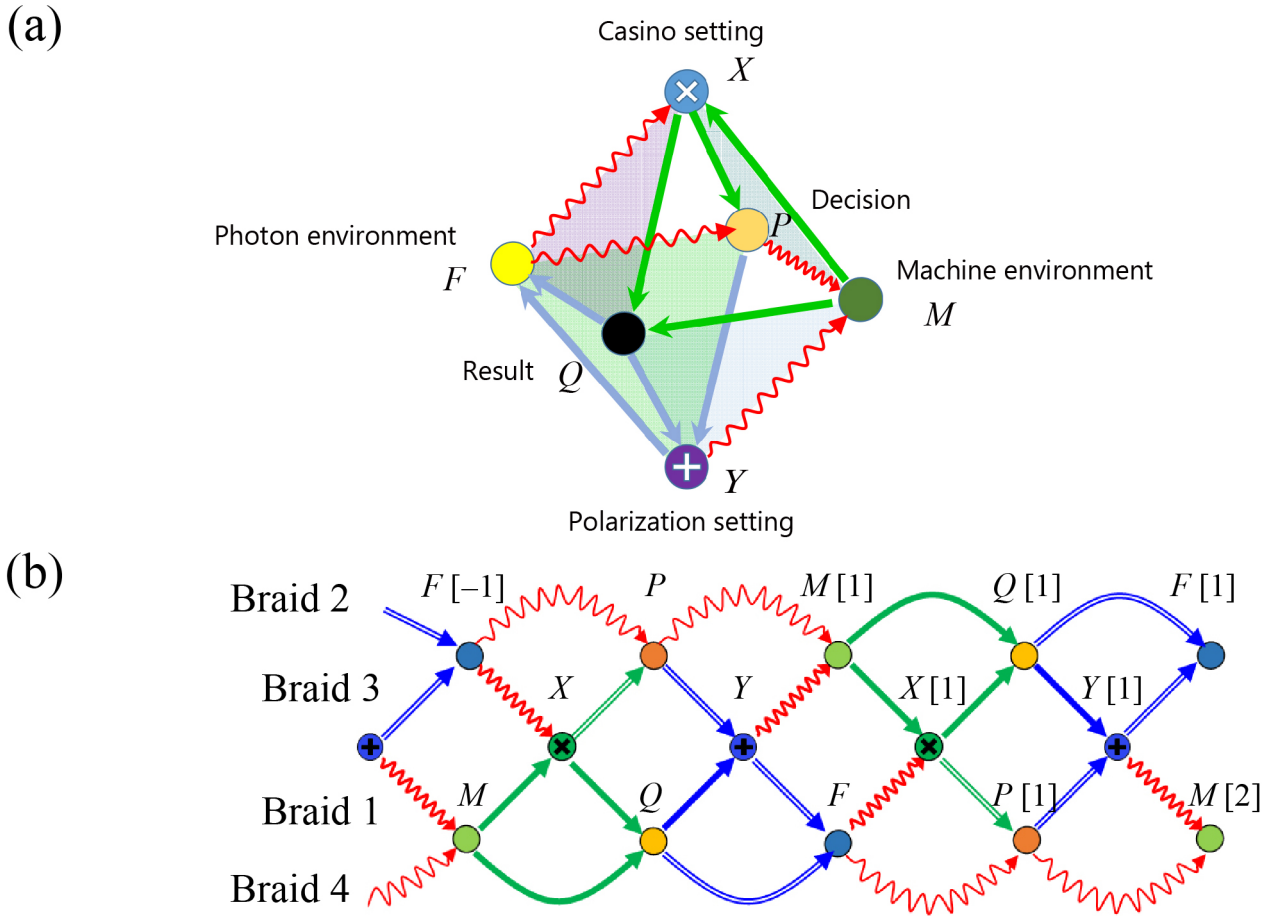

Fig. 6. Category theoretic analysis of photon-based decision making. (b) Total six objects are interacted with each other (shown by arrows) called octahedron structure. (b) The time-domain illustration of the relation between objects in octahedron structure where four braids (Braid 1, 2, 3, and 4) are interacting with each other called braid structures.