CONTROLLING CHAOTIC ITINERANCY IN LASER DYNAMICS FOR REINFORCEMENT LEARNING

RYUGO IWAMI1*, TAKATOMO MIHANA1, KAZUTAKA KANNO1, SATOSHI SUNADA2,3, MAKOTO NARUSE4, ATSUSHI UCHIDA1*

Photonic artificial intelligence has attracted considerable interest in accelerating machine learning; however, the unique optical properties have not been fully used for achieving higher-order functionalities. Chaotic itinerancy, with its spontaneous transient dynamics among multiple quasi-attractors, can be used to realize brain-like functionalities. In this study, we numerically and experimentally investigate a method for controlling the chaotic itinerancy in a multimode semiconductor laser to solve a machine learning task, namely, the multiarmed bandit problem, which is fundamental to reinforcement learning. The proposed method uses chaotic itinerant motion in mode competition dynamics controlled via optical injection. We found that the exploration mechanism is completely different from a conventional searching algorithm and is highly scalable, outperforming the conventional approaches for large-scale bandit problems. This study paves the way to use chaotic itinerancy for effectively solving complex machine learning tasks as photonic hardware accelerators.

INTRODUCTION

Photonic accelerators provide fast and efficient information processing by using photonic technologies to overcome the limitations of integrated circuit density in semiconductor technologies, known as the end of Moore’s law (1, 2). Notably, photonic accelerators have been discussed for certain dedicated information processing in machine learning, such as deep neural network, and they are expected to contribute to artificial intelligence (AI). Photonic accelerators can be considered as preprocessors that use optical signals combined with electronic computing (1). Recently, principles and technologies of photonic accelerators have been proposed and demonstrated, such as artificial photonic neural networks (3), coherent Ising machines (4), optical pass gate logic (5), photonic reservoir computing (6–8), and photonic decision-making for reinforcement learning (9–13).

Reinforcement learning is a method in which an agent learns actions to maximize the reward from interactions with dynamical environments by trial and error (14). Reinforcement learning differs from supervised learning, in which an agent learns from a training set of correct actions that are provided by a supervisor to adapt to the situations that are not included in the training set. In reinforcement learning, an agent must spontaneously learn which action is the best choice from the interactions with different environments without a training set (14). Reinforcement learning has been widely used to acquire a superior performance in the game of Go (15, 16), signal transmission in elastic optical networks (17), and robot control (18).

Solving the multiarmed bandit problem (14, 19) is crucial to photonic decision-making. The objective of this problem is to maximize the total reward from multiple choices or slot machines, whose hit probabilities are unknown. The multiarmed bandit problem is fundamental to reinforcement learning (14, 20), where an agent has to obtain higher rewards in initially unknown environments, and covers a wide range of applications, including Monte Carlo tree search (21) and channel selections in wireless communications (22). The multiarmed bandit problem addresses one of the most critical challenges in reinforcement learning, i.e., the exploration-exploitation dilemma in maximizing the total rewards. The agent must explore to identify the higher-reward machine and simultaneously exploit a particular machine (14). Sufficient exploration may aid in accurately identifying the best machine. However, an excessive exploration results in a large loss, while an insufficient exploration causes the agent to miss the best machine and prohibits them from gaining more rewards. The selection of the slot machine with the highest hit probability has been successfully achieved using photonic dynamical systems (9–13). The versatile physical properties of optical processes have been used for photonic decision-making, chaotically oscillating temporal waveforms of chaotic lasers (9), spontaneous mode switching in a ring-cavity semiconductor laser (11), and lag synchronization of chaos in mutually coupled semiconductor lasers (12).

The scalability of decision-making, that is, how to deal with the increasing number of slot machines or choices, is crucial. Reportedly, hierarchical architecture (10) and laser network dynamics (13) accommodate a large number of slot machines. However, these methods have several limitations; for example, the performance depends on the arrangement of the slot machines in the hierarchical architecture (10), and the scalability is less efficient in laser networks (13). The scaling exponent, which is a metric used to quantify the scalability of decision-making, is limited to a number between 1 and 2 (10, 13), which implies the importance of more efficient decision-making principles to accommodate a large number of choices. In addition, the Lotka-Volterra competition mechanism, which is applicable to ecological systems, has been used for...
decision-making (23). Biological species in an ecosystem grow their population by competing for resources, which may be limited, and this interspecific competition mechanism can be modeled using simple ordinary differential equations, namely, the competitive Lotka-Volterra equations. A logarithmic scaling with respect to the number of slot machines has been demonstrated by using this mechanism for decision-making. In other words, this mechanism outperforms the software algorithms reported in the literature, as well as the abovementioned photonic approaches. However, it should be noted that a large number of plays must be assumed to satisfy the analysis condition, indicating that the decision-making method based on this mechanism suffers from practical difficulties.

Chaotic itinerancy has been reported in many interdisciplinary research fields, such as physiological activity (24–26), recurrent neural networks (27–29), neurorobotics (30–32), coupled map lattice (33), and optical turbulence (34). Chaotic itinerancy is a phenomenon where multiple unstable attractors, called quasi-attractors, coexist, and the variables of dynamical systems move around these quasi-attractors (26, 27). Chaotic itinerancy is considered essential to understand the emergence of spontaneous activities in the brain (27, 33). In addition, chaotic itinerancy has been used to implement associative memory (35). Recently, spontaneous behavioral switching has been designed by using chaotic itinerancy (36). The implementation of chaotic itinerancy through practical engineering platforms for machine learning is a promising and exciting approach for realizing high functionalities in the brain, such as spontaneous creativity and associative memory.

Chaotic itinerancy has been observed in photonic systems (37, 38) as the chaotic mode competition dynamics among multiple longitudinal modes in a multimode semiconductor laser. A variety of nonlinear dynamics and chaotic behaviors in semiconductor lasers with optical feedback and injection have been reported (39–42). Chaotic antiphase dynamics have been experimentally reported in a multimode semiconductor laser with optical feedback (43), where the temporal waveforms of the multimode semiconductor laser intensities are anticorrelated. In the chaotic mode competition dynamics, the mode with the maximum intensity (called the dominant mode) is dynamically changed in a multimode semiconductor laser with optical feedback (44). In addition, the chaotic mode competition dynamics can be controlled by the injection current (44) and an external optical injection (45–47). Thus, the chaotic mode competition dynamics in a multimode semiconductor laser could be a suitable platform for achieving an effective spontaneous searching ability to explore an optimal choice in the case of multiple uncertainties. Although the controllability of deterministic chaotic systems (48) would result in functionalities of chaotic itinerancy, developing a scheme for controlling chaotic itinerancy in reinforcement learning–based applications is a considerable challenge.

In this study, we design and conduct the investigation to assess the feasibility of photonic decision-making by controlling the chaotic itinerancy (i.e., mode competition dynamics) both numerically and experimentally in a multimode semiconductor laser with optical feedback and injection. We solve the multiamplified bandit problem, which is the foundation of reinforcement learning, by using chaotic itinerancy for efficient exploration over many choices. We examine the scalability of the number of choices and demonstrate that the chaotic itinerancy–based method outperforms the upper confidence bound 1 (UCB1)–tuned method, which is one of the most well-known software algorithms. The present study is conducted to investigate chaotic itinerancy to use the unique physical characteristics of laser dynamics, as well as to address the scalability issues of photonic decision-making principles. To the best of our knowledge, this is the first demonstration of using chaotic itinerancy for accelerating reinforcement learning tasks and establishing a concrete photonic hardware architecture comprising technologically feasible device elements.

RESULTS
Multimode semiconductor laser with optical feedback and injection

Figure 1 schematically shows the system architecture and dynamics of a multimode semiconductor laser with optical feedback and injection. Five longitudinal modes of the multimode semiconductor laser are assumed to be excited, whose optical frequencies are denoted as $v_m$ for the $m$th modal intensity ($m = 1, 2, ..., 5, v_i < v_j$ for $i < j$). In addition, a single-mode semiconductor laser with an optical frequency $f_m$ is used for the optical injection. The optical output of the single-mode laser is injected into the $m$th modal intensity with a frequency $v_m$ in the multimode semiconductor laser to control the mode competition dynamics, as shown in Fig. 1. $f_m$ is slightly detuned from $v_m$ to achieve injection locking.

We use a numerical model for a multi–longitudinal mode semiconductor laser with optical feedback (41, 44, 49). This model equation is an extension of the Lang-Kobayashi equations (41, 42, 50), which are well-known numerical model equations for a semiconductor laser with optical feedback. We also added an optical injection term from a single-mode semiconductor laser (51). Modes 1, 2, ..., $M$ are assigned from the lower- to higher-frequency modes. This multimode semiconductor laser system is an autonomous system without optical injection. Details of the numerical model used in this study are provided in Materials and Methods.

Figure 2 summarizes the temporal waveforms of the optical intensity, $I_m(t) = |E_m(t)|^2$, obtained via numerical simulation. Figure 2A shows the temporal waveforms of the modal intensities without optical injection. All the modal intensities oscillate chaotically, and the temporal dynamics of the modal intensities are different. Figure 2B shows the temporal waveform of the total intensity, which is the sum of the modal intensities $I_{total}(t) = \sum I_m(t)$; the total intensity oscillates chaotically. Figure 2C shows the temporal waveforms of the modal intensities when the optical signal is injected.
into mode 3 with an injection strength of $\kappa_{\text{inj},3} = 6.0 \text{ ns}^{-1}$. The oscillation of mode 3, denoted by the red curve, is enhanced by the optical injection, and the duration of the dominant mode for mode 3 is longer than that for the other modes. Therefore, the oscillations of the other modes are suppressed. Figure 2D shows the temporal waveform of the total intensity obtained from Fig. 2C, and this temporal waveform also exhibits chaotic oscillation. Notably, the optical injection does not result in any substantial change in the mean of the total intensity, because the oscillation of mode 3 is enhanced, and those of the other modes are suppressed spontaneously owing to the power conservation in the total intensity [i.e., antiphase dynamics (43)]. Bifurcation diagrams of the total intensity are presented in fig. S1 to show the transition between a steady state and a chaotic oscillation.

Figure 3A shows the temporal waveforms of the optical frequency detuning for the total intensity and the five modal intensities without optical injection. The optical frequency detuning of the total and modal intensities from the central frequency is calculated from the phase equations. The optical frequency detuning of the total intensity (black curve) dynamically moves among the different modes, and the transition of the modes occurs spontaneously. Here, we clearly observe chaotic itinerancy among the five modes that are unstable quasi-attractors. The total intensity stays at one of the modes and shifts to the neighboring modes with time. The optical frequency detuning of the total intensity fluctuates very rapidly, corresponding to the longitudinal mode spacing $\Delta \nu = 35.5 \text{ GHz}$. Figure 3B shows the temporal waveforms of the optical frequency detuning for the total intensity and the five modal intensities with optical injection into mode 3 ($\kappa_{\text{inj},3} = 6.0 \text{ ns}^{-1}$). Although chaotic itinerancy is still observed, the total mode remains at mode 3 for a relatively long duration. The residence time of the total intensity in other modes is reduced as shown in Fig. 3B, compared with that shown in Fig. 3A, owing to the optical injection to mode 3. In this manner, the chaotic itinerancy can be controlled via optical injection (see movies on chaotic itinerancy among the five modes and the corresponding quasi-attractor transition in the phase space with and without optical injection in movies S1 and S2).

We investigated the residence time of the total intensity on one of the modes when chaotic itinerancy occurred without optical injection. The residence time is defined as the duration of the optical frequency of the total intensity (obtained from the phase of the total electric-field amplitude) staying in one mode over one oscillation period, which is 28 ps, determined by a longitudinal mode spacing of 35.5 GHz. We measured the residence time for each modal intensity and created a histogram of the residence time to evaluate the ratio of observing certain residence times, followed by evaluating the probability of residence time. Figure 3C shows the probability of the residence time of the total intensity for each modal intensity without optical injection; these probabilities are calculated from 10-ms-long temporal waveforms. The curves exhibit a linear reduction as a function of residence time on a semi-logarithmic scale. Therefore, we found an exponential relationship of the residence time probability as $P = Ae^{\beta t}$, where $t$ denotes the residence time, and $A$ and $\beta$ are real numbers. The exponents of the exponential decay $\beta$ are different for the five modes: $\beta = -2.7$, $-2.0$, $-1.6$, $-2.0$, and $-2.7$ for modes 1, 2, 3, 4, and 5, respectively. The mean residence times $\tau_r = 1/|\beta|$ are 0.37, 0.50, 0.63, 0.50, and

![Fig. 2. Temporal waveforms of multimode semiconductor laser with optical feedback.](image-url)
0.37 ns for modes 1, 2, 3, 4, and 5, respectively. \( \tau \) is the maximum at the central mode (i.e., mode 3), whereas it decreases as the mode becomes closer to the mode at the edge (modes 1 and 5). Therefore, laser dynamics are highly likely to provide a relatively stable residence when the mode is located at the center, whereas it explores other modes when the mode is located far from the central mode.

Figure 3D shows the probability of the residence time of the total intensity for each modal intensity under optical injection in mode 3. The residence time in mode 3 is enhanced via optical injection, and the absolute value of the slope of the probability curve decreases (indicated by the red curve). In contrast, the residence time in other modes is reduced, and the absolute value of the slope is increased. It is worth noting that different slopes are observed in the regions of short (<1 ns) and long (>1 ns) residence times for all the modes. Therefore, the statistical characteristics of the chaotic itinerancy can be altered via optical injection.

We measured the change in the dominant mode ratio under optical injection. The dominant mode ratio represents the ratio of the probability at which the \( m \)th modal intensity reaches the maximum value over time. The dominant mode ratio \( \text{DMR}_m \) of mode \( m \) is defined as follows (44)

\[
\text{DMR}_m = \frac{1}{d} \sum_{l=1}^{d} D_m(l)
\]

where \( d \) is the total number of sampling points and corresponds to the length of the temporal waveform used for the calculation. \( D_m(l) \) is the function that returns 1 if mode \( m \) has the maximum value among the other modes at the \( l \)th sampling point, and 0 otherwise.

Figure 4A shows the dominant mode ratio as a function of the optical injection strength subjected to mode 3, ranging from 0.0 to 15.0 ns\(^{-1}\). The dominant mode ratio is calculated from the temporal waveforms over 20 μs for each optical injection strength. In Fig. 4A, the dominant mode ratio of mode 3 (red line) drastically increases when the optical injection strength becomes greater than approximately 2.5 ns\(^{-1}\). When the optical injection strength is greater than approximately 8.0 ns\(^{-1}\), the dominant mode ratio of mode 3 is one, meaning that mode 3 becomes the perfectly dominant mode.

---

**Figure 3.** Chaotic itinerancy of total intensity among five modes with different oscillation frequencies. (A and B) Temporal waveforms of the optical frequency detuning of the total and modal intensities, and (C and D) probabilities of the residence time of total intensity in each mode. (A and C) No optical injection and (B and D) with optical injection to mode 3 (injection strength \( k_{\text{inj},3} = 6.0 \text{ ns}^{-1} \)).

**Figure 4.** Dominant mode ratio for five modes as a function of the optical injection strength. Optical signal is injected into (A) mode 3 and (B) mode 1. The dominant mode ratio is calculated from the temporal waveforms over 20 μs for each optical injection strength.
Figure 4B shows the dominant mode ratio as a function of the optical injection strength subjected to mode 1 (outermost mode). It is worth noting that the initial optical frequency detuning between mode 1 and the optical injection signal must be adjusted carefully to match the optical feedback phase (see Materials and Methods). Under this condition, the enhancement of the mode excitation (i.e., dominant mode control based on optical injection) is effective even for the side modes with a smaller gain. From such characterizations, the probability of a certain mode being the dominant mode can be configured by changing the optical injection strength. In other words, we found that the mode competition dynamics can be controlled by engineering the optical injection into particular modes.

One of the most important parameters to control chaotic itinerancy is the initial optical frequency detuning between single- and multimode semiconductor lasers. The initial optical frequency detuning changes the characteristics of the dominant mode ratio with respect to the optical injection strength and strongly affects the decision-making performance. The change in the dominant mode ratio has a large impact on decision-making, because it determines the transition from exploration to exploitation, which will be described in the next section.

**Decision-making using multimode semiconductor laser with optical feedback and injection**

Here, we describe the decision-making principle for solving the multiarmed bandit problem by using the aforementioned multimode semiconductor laser dynamics. Figure 5 shows a schematic diagram for photonic decision-making using a multimode semiconductor laser with optical feedback and injection. We solve the multiarmed bandit problem with M slot machines by assigning each modal intensity of M longitudinal modes of the multimode semiconductor laser to each slot machine, as shown in Fig. 5. The dominant mode is determined by comparing the modal intensities at a sampling interval, and the slot machine corresponding to the dominant mode is selected. After the slot machine is selected, the optical injection strength is controlled using the tug-of-war method (52–55), according to the result of the slot machine selection. We assume that the result of the slot machine selection is "hit" or "miss" with the hit probability \( P_m \) (miss probability: \( 1 - P_m \)) for slot machine \( m \). The rewards of hit and miss correspond to 1 and 0, respectively (binary rewards; see Materials and Methods for details of the decision-making algorithm).

We performed decision-making by controlling the optical injection strength of single-mode semiconductor lasers. The mode corresponding to a large evaluation value \( X_m(p) \) (see Materials and Methods) can be enhanced via optical injection, because the slot machine with a large \( X_m(p) \) is evaluated as a good slot machine. The optical injection strength for mode \( m \) is controlled at the \( pt \)th play, as follows

\[
\kappa_{inj,m} = \begin{cases} 
\kappa_{inj,max} & \text{if } (kX_m(p) > \kappa_{inj,max}) \\
(kX_m(p))(0 \leq kX_m(p) \leq \kappa_{inj,max}) & \text{if } \kappa_{inj,max}(kX_m(p) < 0) \\
0 & \text{otherwise} 
\end{cases}
\]

where \( k \) is the coefficient that adjusts the optical injection strength, and \( \kappa_{inj,max} \) is the upper limit of the optical injection strength. In this study, we set \( k = 0.1 \text{ ns}^{-1} \) and \( \kappa_{inj,max} = 15.0 \text{ ns}^{-1} \).

We examined the decision-making performance when the number of slot machines \( M \) was changed. In this case, the modal intensities are sampled at a sampling interval of 0.1 ns and subsequently compared. Thus, the slot machine is selected every 0.1 ns. The hit probability of slot machine 1 is set to 0.9, and those of the other slot machines are set to 0.7. The slot machine is assigned to mode \( i \) in the multimode laser with \( M \) modes. Therefore, the slot machine with the maximum hit probability is assigned to the outermost mode (mode 1). Figure 6A shows the results of slot machine selection as the number of plays is increased for \( M = 5 \). All the slot machines are selected almost randomly when the number of plays is small. Slot machine 1 is selected more frequently as the number of plays increases. After approximately the 800th play, only slot machine 1 is selected. Therefore, an accurate slot machine selection is accomplished. Figure 6B depicts the case in which the number of slot machines is large, \( M = 129 \). Evidently, in the initial phase, the slot machines corresponding to the modes near the central frequency are extensively selected. The selection range is then expanded to other modes as the number of plays increases. Slot machine 1, which is the highest hit probability machine, can be selected successfully after approximately the 6000th play (movies of the decision-making process using the mode competition dynamics for \( M = 5 \) and 129 are shown in movies S3 and S4).

Furthermore, we investigated the statistical characteristics of decision-making performance. As shown in Fig. 6A, consecutive plays are conducted and repeated every 500 cycles to evaluate a statistical measure of the correct decision rate (CDR) (9–13), which is defined as follows

\[
\text{CDR}(p) = \frac{1}{S} \sum_{l=1}^{S} C(l, p) 
\]

where \( S \) is the total number of cycles; \( C(l, p) \) is a function that returns 1 if the slot machine with the highest hit probability is selected at the \( pt \)th play and the \( lh \) cycle, and 0 otherwise; and \( \text{CDR}(p) \) represents the average rate in \( S \) cycles, i.e., the rate at which the slot machine with the highest hit probability is selected at the \( pt \)th play.

---

**Fig. 5. Schematic of the decision-making method using a multimode semiconductor laser with optical feedback and injection.** \( v_m \) is the frequency of the \( m \)th longitudinal mode of the multimode semiconductor laser, and \( f_m \) is the frequency of the single-mode semiconductor laser for optical injection to the \( m \)th mode. BS, beam splitter; ISO, optical isolator; VA, variable attenuator.
Figure 6C shows the CDR as the number of plays is changed when the number of slot machines \( M \) ranges from 3 to 513. More precisely, \( M \) is specified by \( 2^n + 1 \), where \( n \) is a positive integer ranging from 1 to 9. The hit probability of slot machine 1 is set to 0.9, and those of the other slot machines are set to 0.7 for all \( M \). The CDR exhibits a smaller value when the number of plays is small and gradually increases as the number of plays increases. The CDR approaches one as the number of plays increases for all \( M \), as observed in Fig. 6C, implying that the highest hit probability slot machine is accurately selected. In addition, the curves shown in Fig. 6C are almost equidistant on the horizontal logarithmic scale as the number of slot machines \( M \) increases in the form of \( 2^n + 1 \). This result indicates that the scaling law between the number of plays for correct decisions and the number of slot machines can be obtained from these curves.

Furthermore, we investigated a well-known quantitative measure of regret to evaluate the decision-making performance. Regret at the \( p \)th play is defined as the difference between the ideal (maximum) total reward and the actual reward obtained by the plays, as follows (56, 57):

\[
\text{Regret}(p) = p \cdot P_{\text{max}} - \frac{1}{S} \sum_{l=1}^{S} \sum_{m=1}^{M} [P_{m} S_{l,m}(p)] \tag{4}
\]

where \( P_{\text{max}} \) is the maximum hit probability, \( S \) is the total number of cycles, \( P_{m} \) is the hit probability of slot machine \( m \), and \( S_{l,m}(p) \) is the number of selections for slot machine \( m \) at the \( l \)th cycle until the \( p \)th play. A smaller regret indicates a better decision-making performance. In this study, we set \( S = 500 \) to evaluate the regret. Figure 6D shows the regret as the number of plays changes when the number of slot machines \( M \) ranges from 3 to 513. The regret curves increase with an increase in the plays and saturate at certain values for all \( M \). The saturation of the curves indicates that correct decision-making is achieved after a sufficient number of plays are conducted. It is also worth noting that the curves are equidistantly distributed on a logarithmic scale; hence, a scaling law can be obtained from these curves.

**Scalability of decision-making performance**

We investigated the scalability of the decision-making performance when the number of slot machines is changed. We analyzed the number of plays when the CDR reaches 0.95, as shown in Fig. 6C, to examine scalability (10, 13). The red curve in Fig. 7A shows the scalability of the number of plays required for the CDR of 0.95.
Iwa mi

and obtain Regret = 33.7 (denoted as chines, whereas the scaling exponents for the previous methods are between 1 and 2 (10, 13). We compared the scalability of our scheme with that of the UCB1-tuned method (56), which is a well-known software algorithm for solving the multiarmed bandit problem (see Materials and Methods for details). The blue curve shown in Fig. 7A indicates the number of plays \( N_{\text{play}} \) required to achieve a CDR of 0.95 as the number of slot machines \( M \) increases for the UCB1-tuned method. When the number of slot machines is small, the UCB1-tuned method shows a superior performance because it requires fewer plays for correct decision-making. However, when the number of slot machines increases, our method is superior in achieving the correct decision-making with fewer plays. We approximate these curves by using the power law. The curve for the UCB1-tuned method is approximately given by \( N_{\text{play}} = 93M^{1.06} \), that is, the scaling exponent of the power law for the UCB1-tuned method is 1.06, which is larger than that for the proposed method of 0.70. The difference in the scaling exponents can influence the search capability of the best choice, particularly for an extremely large-scale multiarmed bandit problem. For example, the proposed method enables us to determine the best choice at a 2.7 times faster rate, compared to the UCB1-tuned method for \( M = 500 \). The best choice can be determined 6.3 times faster for \( M = 5000 \).

We also obtained the scaling exponent from the results of regret, as shown in Fig. 6D. Figure 7B shows the regret at the number of plays \( p = 60,000 \) as the number of slot machines \( M \) increases for the multimode laser (red) and UCB1-tuned method (blue). We approximate the curves shown in Fig. 7B by a power law \( \text{Regret} = b \cdot M^\gamma \) and obtain \( \Gamma = 0.73 \) and 1.06 for the multimode laser and UCB1-tuned method, respectively. Therefore, the scaling exponent for the multimode laser is smaller than that for the UCB1-tuned method, even when using the regret measure. A better decision-making performance can be achieved using the multimode laser for more than 100 slot machines, as shown in Fig. 7B. The scaling results for the different settings of the hit probabilities and the assignment of the slot machines to the modes are summarized in figs. S2 and S3. We also investigate the robustness of the decision-making performance against spontaneous emission noise in the multimode laser, as shown in fig. S4.

To understand the improvement in the performance of our method, we compared the characteristics of the slot machine selection between the proposed method and the UCB1-tuned method. We calculated the Shannon entropy from the probabilities of the number of selections for each slot machine to evaluate the bias of the selection probabilities (see Materials and Methods for details of the calculation of the Shannon entropy) (58). A bias in the slot machine selection exists if the entropy is small. In particular, only one slot machine is always selected if the entropy is 0. Figure 8A shows the Shannon entropy of the probabilities of the number of selection for each slot machine when the number of slot machines is \( M = 5 \), as the number of plays changes, using the multimode laser and UCB1-tuned method. The entropy for the UCB1-tuned method (blue curve in Fig. 8A) decreases gradually and almost monotonically as the number of plays increases on a double logarithmic scale. In contrast, the entropy for the multimode laser (red curve in Fig. 8A) is almost constant when the number of plays is small, and it suddenly decreases after approximately the 600th play. In other words, the multimode laser system explores a variety of selections at the beginning, and then, accurate decision-making is suddenly accomplished, as demonstrated in Fig. 6A. We interpret that the constant entropy corresponds to the exploration procedure, where all the slot machines are selected equally, whereas the sudden dropout of the entropy corresponds to the exploitation procedure, where the slot machine selection is suddenly "accelerated" after sufficient exploration owing to the chaotic mode competition dynamics.

Figure 8B shows the entropy when the number of slot machines is increased to \( M = 129 \). The entropy of the UCB1-tuned method also decreases gradually as the number of plays increases, similar to the case of \( M = 5 \) in Fig. 8A. Similarly, the entropy for the multimode laser is constant at the beginning; it slightly increases, which coincides with the expansion of the selection range, as observed in Fig. 6B. The entropy decreases rapidly after approximately the 5000th play, realizing a low entropy much faster than that achieved using the UCB1-tuned method. On the basis of the results shown in Fig. 8B, we consider that the exploration process is similar for the multimode laser system and UCB1-tuned method. However, the exploitation is accelerated by the multimode laser method. Therefore, the proposed method using a multimode laser outperforms the UCB1-tuned software algorithm. The multimode laser can accelerate the process of slot machine selection when a large number of slot machines exist owing to the fast convergence of the chaotic mode competition dynamics (see the comprehensive investigation of Shannon entropy for different numbers of slot machines in fig. S5).

Experimental implementation

In the previous sections, we explained our numerical results of decision-making. In this section, we describe our experimental implementation of the proposed method using a semiconductor laser with optical feedback. We performed an online decision-making experiment in a feedback-loop configuration for searching the slot
machines and controlling the mode competition dynamics. Furthermore, fiber-based optical components were used in the experiment, as described in fig. S6 (see the Supplementary Materials for details).

Figure 9 shows the experimental results of the mode competition dynamics and decision-making realized using the multimode semiconductor laser. Figure 9A depicts the temporal waveforms of the four-longitudinal mode dynamics, when the normalized optical injection strength for each mode is set to 0.05. Figure 9B presents the temporal waveforms of the four modes, when the normalized optical injection strength for mode 1 is increased to 1.0, while those for the other modes remain 0.05. Figure 9C shows the dominant mode ratio, when the normalized optical injection strength for mode 1 is changed. Figure 9D shows the experimental result of the CDR, when the decision-making with four slot machines is performed for 100 cycles. Slot machine 1 is assigned to the highest hit probability, and the hit probabilities are set to \((P_1, P_{2,3,4}) = (0.9, 0.1), (0.8, 0.2),\) and \((0.7, 0.3)\) for three different experiments, respectively. All the cases, the CDR converges to 1 as the number of plays increases, and the correct decision-making is achieved experimentally. A faster convergence to the CDR of 0.95 is realized, when a larger difference in the hit probabilities is used, because it is easier to find the slot machine with the highest hit probability with less exploration. The experimental results agree well with those of our numerical simulations, validating the successful experimental implementation of the proposed method.

**DISCUSSION**

Our results show that chaotic itinerancy (i.e., the mode competition dynamics) in a multimode semiconductor laser provides efficient exploration and exploitation to identify the slot machine with the largest reward (i.e., best choice). This partly originates from the chaotic “partition” of lasing energy, which is highly sensitive to external stimuli (e.g., optical injection) [41]. Our results indicate that optical injection can result in an efficient energy concentration in a particular mode corresponding to the best choice, even when many
modes are involved in the lasing state. In other words, one of the multiple modes can be easily enhanced via optical injection, whereas the other modes are suppressed, because the oscillation of each modal intensity is very weak in the presence of a large number of modes. The convergence of the total intensity to one of the modes, i.e., a quick transition from the exploration to exploitation phase, can be accelerated even for a low optical injection energy, which suggests the high feasibility of the implementation in real hardware.

Our method using multimode laser dynamics outperforms the UCB1-tuned algorithm when the number of slot machines is very large (more than 100). The UCB1-tuned algorithm selects slot machines in parallel based on the confidence bound, which gradually decreases the entropy; however, acceleration cannot be induced. Therefore, our method based on multimode laser dynamics can select the correct slot machine much faster than the UCB1-tuned algorithm when the number of slot machines is large. The scaling exponent of the proposed chaotic itinerary–based method is 0.70. This indicates the advantage of the proposed method under a large number of slot machines, compared to the existing software algorithms and other photonic methods, whose scaling exponents are summarized in Table 1. The exponent is 1.06 for the UCB1-tuned algorithm, while those for the photonic methods reported in (10) and (13) are 1.16 and 1.85, respectively. The identification of the best choice from many choices with unknown rewards is crucial in practical applications such as continuous problems, including online auction or routing (59) or channel searching in wireless and optical communications (17, 60). The proposed photonic approach may open a pathway for solving such large-scale bandit problems.

This scheme uses the combination of a physical competition mechanism in multimode laser dynamics and software control based on the tug-of-war method. Thus, improving the software control algorithm could improve the decision-making performance, including the hit probability estimation (see Materials and Methods). In addition, the mode competition dynamics could be tuned by changing the laser parameters, such as the injection current, feedback strength, and feedback delay time, to optimize the decision-making performance.

Although the speed of slot machine selection and reward reception will limit the decision-making performance, and the advantages of photonic decision-making in terms of speed are not fully used in the current situation, various potential applications of fast decision-making can still be expected. For example, in the current stock trading, the decision is made in the order of milliseconds using AI; this process is known as FinTech (61). However, in the near future, the speed of the automated trading systems can be increased via fast photonic decision-making. Furthermore, the next-generation optical communication systems require complex functionalities, such as automatic selection of communication channels, symbol formats, and routing control by AI (17), which can also be realized via fast photonic decision-making. Thus, fast photonic decision-making is a promising approach for such future applications.

Our decision-making method could be applied to other nonlinear dynamical systems that produce chaotic itinerary. The spontaneous searching ability supported by chaotic itinerary is extremely promising for solving complex machine learning tasks, as well as for understanding the spontaneous activities of the brain. The engineering design of chaotic itinerary is crucial to maximize the searching ability using control techniques (36). The combination of chaotic itinerary and control techniques results in a research direction for machine learning applications.

In this study, we successfully performed photonic decision-making by controlling the chaotic itinerary, known as the mode competition dynamics, in a multi–longitudinal mode semiconductor laser with optical feedback and injection to solve the multiarmed bandit problem both numerically and experimentally. We confirmed that the mode competition dynamics in a multimode semiconductor laser can be controlled via optical injection in a specific mode. We assigned one slot machine to each modal intensity and controlled mode competition dynamics, based on the results of the slot machine selection. We solved the multiarmed bandit problem with up to 513 slot machines and evaluated the decision-making performance using the CDR and regret. We found that our decision-making scheme shows excellent scalability between the number of plays for correct decision-making and the number of slot machines in the form of a power law with a scaling exponent of 0.70, which is superior to that of the well-known UCB1-tuned software algorithm. We investigated the entropy of selection probabilities and found that our decision-making scheme provides fast and efficient decision-making for the correct slot machine when the number of slot machines is larger than 100. Our method using multimode laser dynamics can enhance the decision-making performance under a large number of choices.

To conclude, this study demonstrated that chaotic itinerary in multimode laser dynamics is a promising resource for solving machine learning tasks as photonic accelerators. The proposed chaotic itinerary–based principle exploits the high-bandwidth attributes of light as well as complex laser dynamics, which are manifested by the residence time statistics and entropy analysis. The physical properties and architectural design allow an efficient and accelerated bandit problem solving. On the basis of the insights gained through the present study, the proposed method that combines chaotic itinerary and complex laser dynamics can be extended to solve higher-order problems and complex machine learning tasks in the future.

Table 1. Comparison of the scaling exponents obtained from the graphs for the number of plays required for the CDR of 0.95 as a function of the number of slot machines.

| Method                      | Scaling exponent |
|-----------------------------|------------------|
| Multimode semiconductor laser (proposed) | 0.70             |
| UCB1-tuned algorithm (56)    | 1.06             |
| Chaotic temporal waveforms (10)| 1.16             |
| Laser network (13)          | 1.85             |

MATERIALS AND METHODS

Numerical model of multimode semiconductor laser

The numerical model of the multimode semiconductor laser with $M$ longitudinal modes is described by the following deterministic...
equations (44, 49)

$$\frac{dE_m(t)}{dt} = \frac{1 - i\alpha}{2} \left\{ \frac{G_m[N(t) - N_0]}{1 + \varepsilon \sum_{u=1}^{M} |E_u(t)|^2} - \frac{1}{\tau_p} \right\} E_m(t)$$

$$+ \kappa E_m(t - \tau) \exp(i\omega_m \tau)$$

$$+ \kappa_{inj,m} A_s \exp(-i2\pi \Delta f_{inj} t)$$

(5)

$$\frac{dN(t)}{dt} = J - \frac{N(t)}{\tau_s} - \sum_{v=1}^{M} \left\{ \frac{G_v[N(t) - N_0] |E_v(t)|^2}{1 + \varepsilon \sum_{u=1}^{M} |E_u(t)|^2} \right\}$$

$$G_m = G_N \left[ 1 - \frac{(v_m - v_m)^2}{\Delta \nu_g^2} \right]$$

(6)

where $E_m(t)$ represents the slowly varying complex electric-field amplitude of the $m$th longitudinal mode and $N(t)$ represents the carrier density. In Eqs. 5 and 6, $\alpha$ is the linewidth enhancement factor, $G_m$ is the gain coefficient of the $m$th mode, $N_0$ is the carrier density at transparency, $\varepsilon$ is the gain saturation coefficient, $\tau_p$ is the lifetime of the photon, $\tau$ is the carrier lifetime, $\kappa$ is the optical feedback strength of the multimode semiconductor laser, $\kappa_{inj,m}$ is the optical injection strength from the single-mode semiconductor laser to the $m$th mode of the multimode semiconductor laser, $\tau$ is the round-trip time of light in the external cavity of the multimode laser, and $\omega_m$ is the angular frequency of the $m$th mode ($\omega_m = 2\pi v_m$). Here, $A_s$ represents the steady-state solution of the electric-field amplitude of a single-mode semiconductor laser without optical feedback and injection (51). $\Delta f_m$ is the initial optical frequency detuning between the single-mode semiconductor laser with frequency $f_m$ and the $m$th mode of the multimode laser ($\Delta f_m = f_m - v_m$), and $J$ is the injection current. The gain coefficient of the $m$th mode is approximated by a parabolic gain profile, as shown in Eq. 7, where $G_N$ is the gain coefficient at the central mode $m_c = (M + 1)/2$ and $M$ is assumed to be an odd integer. $\Delta \nu_g$ is the frequency width of the gain profile. We define the frequency of the $m$th mode as $v_m = v_{mc} + (m - m_c)\Delta \nu$, where $v_{mc}$ is the frequency of the central mode $m_c$, and $\Delta \nu = 35.5$ GHz is the frequency spacing among the longitudinal modes. In this model, no spontaneous emission noise is included to investigate the dynamics of deterministic chaotic itinerancy, as shown in Fig. S4. The parameter values used in this study are listed in Table S1.

The frequency width of the gain profile is set to $\Delta \nu_g = 141.9$ THz (which corresponds to the wavelength width of 1000 nm) to investigate the scaling properties for a large number of modes and slot machines. Here, $\Delta \nu_g = 141.9$ THz may be too wide for the gain width of an actual semiconductor laser [the typical gain width of semiconductor lasers is several tens of terahertz (or 10 to 100 nm) (44, 49, 62)]. It is necessary to set a large gain width to increase the number of modes (up to 513) that exhibit chaotic mode competition dynamics via optical feedback without the noise term. This is because the modes near the central frequency of the gain curve are raised if the noise term is not included, and the mode coupling is included only via the carrier density (49, 63). We confirmed that similar results described in the main text are obtained when using a smaller gain width, $\Delta \nu_g = 12.70$ THz (wavelength width of 100 nm), and a smaller number of modes as realistic parameter values.

**Tuning of frequency detuning of optical injection**

To perform decision-making, the initial optical frequency detuning between each mode and the optical injection signal must be adjusted. The difference in the optical feedback phase for different longitudinal modes changes the dominant mode ratio. Moreover, the optical injection strength required for obtaining a large dominant mode ratio depends on the optical phases of the feedback and injection signals (64). Considering the mode spacing $\Delta \nu$ and the difference in the optical feedback phase among the modes $2\pi \Delta \nu t$, the initial optical frequency detuning $\Delta f_m$ can be modified with the phase shift among different modes $\Phi_{adj,m}$ as follows

$$\Phi_{adj,m} = 2\pi (m_c - m) \Delta \nu t \pmod{2\pi}$$

(8)

$$\Delta f_m = \Delta f_m - \frac{1}{2\pi \Phi_{adj,m}}$$

(9)

Equation 8 represents the phase shift required to match the optical feedback phase of the $m$th mode with that of the central mode $m_c$. In Eq. 9, the initial optical frequency detuning $\Delta f_m$ can be adjusted to compensate for the phase shift calculated using Eq. 8. We fix the initial frequency detuning for the central mode at $\Delta f_{mc} = -4.0$ GHz, and the $\Delta f_m$ values of the other modes are adjusted using Eqs. 8 and 9 to obtain similar characteristics of the optical injection among the different modes, as shown in Fig. 4 (A and B).

**Decision-making algorithm**

We describe the tug-of-war method for $M$ slot machines used in this study (9–12, 52, 53, 55). The evaluation value $X_m(p)$ for slot machine $m$ at the $p$th play originates from the displacement of the amoeba branch in the tug-of-war method; the initial value is $X_m(0) = 0$. If the slot machine $s$ is played at the $p$th play and the result is hit, $X_m(p)$ is changed as follows

$$X_m(p) = \begin{cases} X_m(p - 1) + \Delta(p)(m = s) \\ X_m(p - 1) - \frac{\Delta(p)}{M - 1}(m \neq s) \end{cases}$$

(10)

where $\Delta(p)$ represents the change in the hit case. If the slot machine $s$ is played at the $p$th play, and the result is missed, then $X_m(p)$ changes as follows

$$X_m(p) = \begin{cases} X_m(p - 1) - \Omega(p)(m = s) \\ X_m(p - 1) + \frac{\Omega(p)}{M - 1}(m \neq s) \end{cases}$$

(11)

where $\Omega(p)$ represents the amount of change in the miss case. The amounts of changes $\Delta(p)$ and $\Omega(p)$ differ between hit and miss.
because either of them needs to be prioritized based on the parameter settings of the hit probabilities (10). The suboptimal values of $\Delta(p)$ and $\Omega(p)$ are determined by the sum of the two highest hit probability values (54, 55). However, the hit probabilities are unknown initially, and $\Delta(p)$ and $\Omega(p)$ need to be estimated using the results of the slot machine selection as follows (12)

$$\Delta(p) = 2 - [P_{\text{top1}}(p) + P_{\text{top2}}(p)]$$

$$\Omega(p) = P_{\text{top1}}(p) + P_{\text{top2}}(p)$$

where $P_{\text{top1}}(p)$ and $P_{\text{top2}}(p)$ represent the highest and second-highest estimated hit probabilities, respectively, based on the results of the slot machine selection.

In the previous studies (9–13), the hit probability of the slot machine $s$ was estimated by the ratio between the number of hits and the number of total plays for the slot machine $s$. However, this method may result in incorrect estimation at the beginning of the selection process when the number of slot machines is large because $\Delta(p)$ and $\Omega(p)$ only require the two highest estimated hit probabilities. In addition, $\Delta(p)$ and $\Omega(p)$ are estimated from a limited number of the slot machine selection results, and incorrect exploitation may be performed if all the slot machines are not selected.

To overcome this limitation, we modify the estimation method of the hit probabilities at the $p$th play for the $m$th slot machine [$P_m(p)$] as follows

$$P_m(p) = \begin{cases} \frac{R_m(p)}{S_m(p)} + 1 & \text{if } S_m(p) \neq 0 \\ P_{\text{unknown}}(p) & \text{if } S_m(p) = 0 \end{cases}$$

$$P_{\text{unknown}}(p) = P_{\text{top1}}(p)$$

where $R_m(p)$ and $S_m(p)$ represent the number of hits and plays for slot machine $m$ until the $p$th play, respectively. An extremely high estimation of the hit probabilities can be avoided by adding 1 to the denominator in Eq. 14 when the number of plays is small. In addition, the estimated hit probability for the slot machine that has never been selected until the $p$th play is set to $P_{\text{unknown}}(p) = P_{\text{top1}}(p)$, where $P_{\text{top1}}(p)$ is the maximum value of $P_m(p)$ in Eq. 14, to facilitate exploration for all the slot machines. According to Eqs. 14 and 15, the second-highest estimated hit probability becomes the same as the highest estimated hit probability until all the slot machines are selected, resulting in the selection of all the slot machines. In general, $P_{\text{unknown}}(p)$ can be set to a constant value in advance; however, the decision-making performance depends on this constant value. Our method is useful because $P_{\text{unknown}}(p)$ is adaptively changed on the basis of the hit probabilities estimated by the exploration.

**UCB1-tuned algorithm**

The UCB1-tuned method is based on a deterministic algorithm for solving the multiarmed bandit problem. The UCB1-tuned algorithm can be described as follows (14, 56). Each slot machine is selected once for up to $M$ plays ($M$ is the number of slot machines). At the $p$th play ($p > M$), slot machine $m$ with the maximum value of the following UCB$_m$ is selected

$$\text{UCB}_m = \frac{R_m(p)}{S_m(p)} + \sqrt{\frac{\ln p}{S_m(p)}} \min \left[ \frac{1}{4}, \frac{\sigma_m^2(p)}{S_m(p)} + \sqrt{\frac{2\ln p}{S_m(p)}} \right]$$

where $R_m(p)$ and $S_m(p)$ represent the number of hits and plays for slot machine $m$ until the $p$th play, respectively; the function min $(a, b)$ returns a smaller value of $a$ and $b$; and $\sigma_m^2(p)$ represents the sample variance of the reward. In the second term on the right-hand side, $\frac{1}{4}$ represents the upper bound on the variance of the random variable, which follows the Bernoulli distribution (e.g., binary rewards of hit or miss). The first and second terms on the right-hand side represent the estimated hit probability and the correction term that incorporates the confidence interval of the estimated hit probability, respectively.

**Shannon entropy of probabilities of slot machine selection**

We calculate the Shannon entropy from the probabilities of slot machine selection. We calculate the selection probability of slot machine $m$ until the $p$th play using window size $W$ as follows

$$P_{\text{sel,m}}(p) = \frac{1}{W} \sum_{l=p-W+1}^{p} \text{Sel}_m(l)$$

where $\text{Sel}_m(l)$ is a function that returns 1 if slot machine $m$ is selected at the $l$th play, and 0 otherwise. The window size $W$ was set to $5M$, where $M$ is the number of slot machines. The Shannon entropy of the selection probability at the $p$th play is calculated as follows

$$H(p) = -\sum_{l=1}^{M} P_{\text{sel,m}}(l) \log_2[P_{\text{sel,m}}(l)]$$

We calculated the Shannon entropy for each cycle and averaged it over 500 cycles, as shown in Fig. 8.

**Supplementary Materials**

This PDF file includes:

- Figs. S1 to S6
- Table S1

Other Supplementary Material for this manuscript includes the following:

- Movies S1 to S4

**REFERENCES AND NOTES**

1. K. Kitayama, M. Notomi, M. Naruse, K. Inoue, S. Kawakami, A. Uchida, Novel frontier of photonics for data processing—Photonic accelerator. *AIP Photonics* **4**, 090901 (2019).

2. B. J. Shastri, A. N. Tait, T. F. de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, P. R. Prucnal, Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021).

3. Y. Shen, N. C. Harris, S. Skirlo, M. Prabh, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Lanorelle, D. Englund, M. Soljačić, Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).

4. T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbiitsu, O. Tadanaga, H. Takenouchi, K. Aiha, K.-I. Kawarabayashi, K. Inoue, S. Utsunomiya, H. Takesue, A coherent Ising machine for 2000-node optimization problems. *Science* **354**, 603–606 (2016).

5. T. Ishihara, A. Shinya, K. Inoue, K. Nozaki, M. Notomi, An integrated nanophotonic parallel adder. *ACM J. Emerg. Technol. Comput. Syst.* **14**, 1–20 (2018).
Acknowledgments
Funding: This study was supported in part by JSPS KAKENHI (JP19H00868, JP20K15185, JP20H00233, and JP22H05195), JST CREST (JPMJCR17N2), and the Telecommunications Advancement Foundation. Author contributions: All the authors contributed to the development and implementation of the concept. A.U. directed the project. R.I. performed the numerical simulations and experiments. R.I., K.K., and A.U. analyzed the data. R.I., T.M., S.S., M.N., and A.U. contributed to the discussion of the results. R.I., S.S., M.N., and A.U. contributed to manuscript writing. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.