

レーザーカオスを用いた超高速強化学習

Laser Chaos-based Ultrafast Photonic Reinforcement Learning

○成瀬 誠¹、寺島悠太²、内田淳史²、金 成主³ (1. 情通機構、2. 埼玉大、3. 物材機構)

○Makoto Naruse¹, Yuta Terashima², Atsushi Uchida², Song-Ju Kim³

1. NICT, 2. Saitama Univ., 3. NIMS

E-mail: naruse@nict.go.jp

人工知能の重要課題である強化学習とは、動的に変化する環境での意思決定や最適化であり、通信網・電力網等のインフラ瞬時最適化、自動運転等の多くの重要な応用に直結している。強化学習の根幹に、当たり確率の未知なスロットマシンからの獲得報酬を最大化する問題（多本腕バンディット問題）がある。報酬の最大化には有利なマシンを知るための探索が必要になるが、過度な探索は損失であり、一方、性急な判断はよい選択を阻む。すなわち探索と決断に難しいトレードオフが存在する。我々は、この問題を光物質系に固有の性質を用いて物理的に解決する「フォトリック知能」を提案し、これまでに、近接場光及び単一光子を用いた強化学習を実験的に実証した^{1,2)}。これらの研究では、光の微細化とエネルギー効率の極限での性質が生かされており、集積性や省エネ性を活かす IoT 応用等が考え得る。一方、フォトリック知能の実現方式としては様々な可能性があり、特に、光の高帯域性や伝搬高速性など時間領域の光の極限性能での強化学習は、新しい価値を産み出す可能性を秘めている。この場合には、超高速性を活かし、データセンター等での瞬時的最適化や調停（アービトレーション）などの応用が検討し得る。

本研究では、時間領域での光の極限での物理的強化学習の研究の手始めとして、レーザーカオス³⁾と「綱引き (Tung-of-War (TOW))」⁴⁾と呼ばれる物理ダイナミクスに基づいた意思決定原理を組み合わせ、2本腕バンディット問題の解決が可能であることを実験的に確認した。すなわち、当たり確率の高いマシンにシステムが自律的に適応することを実証した。具体的には、半導体レーザーと光遅延系により生成したレーザーカオス光を高速にサンプリングし、TOW 原理によって設定する閾値との大小判定のみで意思決定を行う。実験では、サンプリングレート 100 GSsample/s で合計 10^7 点の光強度データを取得した上で、TOW 原理はオフラインで実装した。システムの全体構造を図 1(a)、光強度の時間系列の例を図 1(b)に示す。2台のスロットマシンの報酬確率を $\{0.2, 0.8\}$, $\{0.4, 0.6\}$ と設定したとき、報酬確率の高いマシンを選択した割合 (Correct Selection Ratio) を図 1(c)に示す。10 ns 毎 (1000 サンプル毎) に強制的に報酬確率を入れ替え環境変化を与えているが、システムは自律的に適応していることが分かる。これにより、レーザーカオスの強化学習への応用が初めて実証された。今後、閾値処理等を物理系で実現する原理やより高度な強化学習課題の物理知能化を検討する。

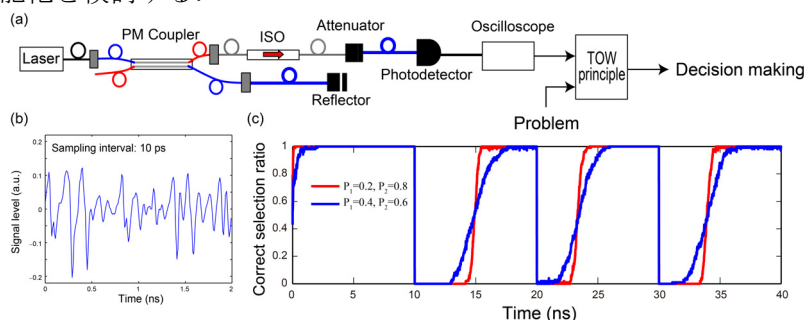


図 1 レーザーカオスを用いた超高速強化学習

謝辞 本研究の一部は日本学術振興会「研究拠点形成事業 (A.先端拠点形成型)」の助成による。

参考文献 1) M. Naruse, et al. J. Appl. Phys. **116**, 154303 (2014). 2) M. Naruse, S.-J. Kim, et al. Sci. Rep. **5** 13253 (2015). 3) 内田淳史, 複雑系フォトリックレーザーカオスの同期と光情報通信への応用— (共立出版, 2016). 4) S.-J. Kim, et al. New J. Phys. **17**, 083023 (2015).