Optimal control for instability suppression in an electrostatic plasma with reinforcement learning

Jinsu Kim
Mechanical and Aerospace Engineering
Princeton University
jk9075@princeton.edu

***Abstract*—This project investigates optimal control of the external electric field to suppress the wave instability, known as the Bump-on-tail instability, in an electrostatic plasma system. We formulate the optimal control problem for instability suppression by minimizing the system's electric field energy and use reinforcement learning to approximate the solution to the Hamilton-Jacobi-Bellman equation. We demonstrate that our approach successfully suppresses the instability, as verified through Fourier analysis and the linear damping rate of the system's electric field.**

***Index Terms*—Vlasov-Poisson equation, Bump-on-tail instability, Hamilton-Jacobi-Bellman equation, Reinforcement Learning**

I. Introduction

Plasma, an ionized gas consisting of ions, electrons, and neutral particles, exhibits complex nonlinear dynamics governed by self-consistent electromagnetic fields [1]. Among the compelling applications of plasma physics is the pursuit of thermonuclear fusion as a future energy. However, the inherent instabilities caused by nonlinear kinetic interactions are severe, since they can grow exponentially through wave-particle interactions [2]. This leads to energy loss and damps the fluctuations. Therefore, it is necessary to investigate how to suppress these instabilities to improve the plasma confinement.

This project investigates suppressing the Bump-on-tail instability in a 1D Vlasov-Poisson plasma by controlling the external electric field through reinforcement learning. The Bump-on-tail instability is an example of the kinetic instabilities induced by a localized perturbation on the electron velocity distribution [2]. This perturbation drives exponential growth of fluctuations and transfers the energy from particles to waves through inverse Landau damping. We simplify to a one-dimensional case with a parameterized external electric field. The cost functional corresponding to our goal is then defined and discretized. Since the nonlinearity of this system makes it difficult to solve the optimal control problem directly, we approximate the value function that satisfies the Hamilton-Jacobi-Bellman (HJB) equation in our problem by reinforcement learning.

This paper is organized as follows. First, we will introduce the Vlasov-Poisson plasma system. Then, we will cover the optimal control problem for instability suppression and the connection between HJB equation and reinforcement learning. The remainder of this paper will present numerical results and discuss the validity of the optimal control in suppressing the instability.

II. The Physical Model

A. Vlasov-Poisson equation

The dynamics of charged particles in a plasma is described by the Vlasov-Maxwell (VM) equation, which governs the time evolution of the distribution in a phase space [3]. For an electrostatic collisionless plasma on a short timescale, the magnetic field is ignored, and ions are stationary. The VM equation is then approximated by the Vlasov-Poisson (VP) equation.

$$\begin{aligned} \frac{\partial f}{\partial t} + v \cdot \nabla_x f - E \cdot \nabla_v f = 0 \\ \nabla^2 \phi(x,t) = \int_{\Omega_v} f dv - 1 \\ E(x,t) = -\nabla \phi(x,t) + E_{in}(x,t) \end{aligned} \tag{1}$$

Here, $f(x,v,t)$ is a distribution function of electrons in a phase space, E and ϕ are the net electric field and its potential, respectively. We define a parameterized external electric field E_{in} as a control input to suppress the instability.

$$E_{in} = \sum_{n=1}^{m} a_n(t) \cos \frac{2\pi n}{L} x + b_n(t) \sin \frac{2\pi n}{L} x \tag{2}$$

where m is the maximum mode number of the input electric field. The corresponding ODE for the motions of electrons is given by 3.

$$\dot{x} = v \quad \dot{v} = -E_{in}(x,t) + \nabla \phi(x,t) \tag{3}$$

B. Numerical discretization

For numerical simulations, we use the Particle-In-Cell method [4], discretizing the system by n super-particles with a particle-grid mapping. First, one can discretize the distribution function from 1 approximated by the superposition of n super-particles of the coordinate (x_i, v_i) and weight w_i.

$$f(x,v,t) \approx \sum_{i=1}^{n} w_i \delta(x - x_i(t)) \delta(v - v_i(t)) \tag{4}$$

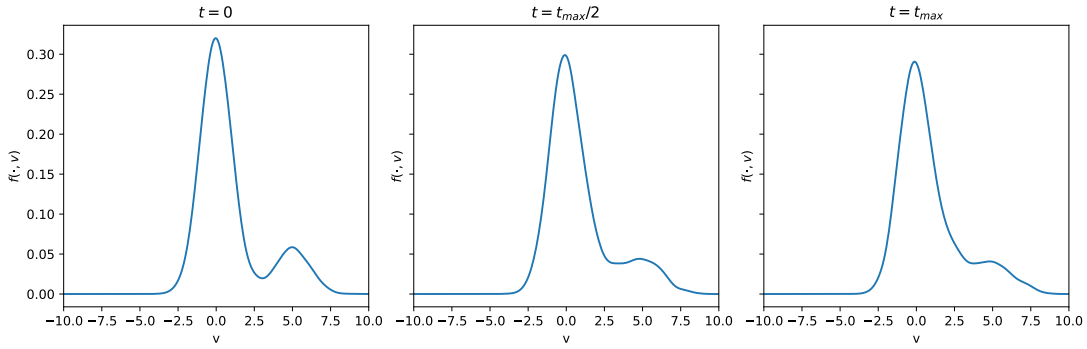

Fig. 1. The electron velocity distribution by the perturbed wave with $A = 0.2$, $n_{mode} = 2$, $v_{th} = 1.0$, and $v_b = 5.0$.

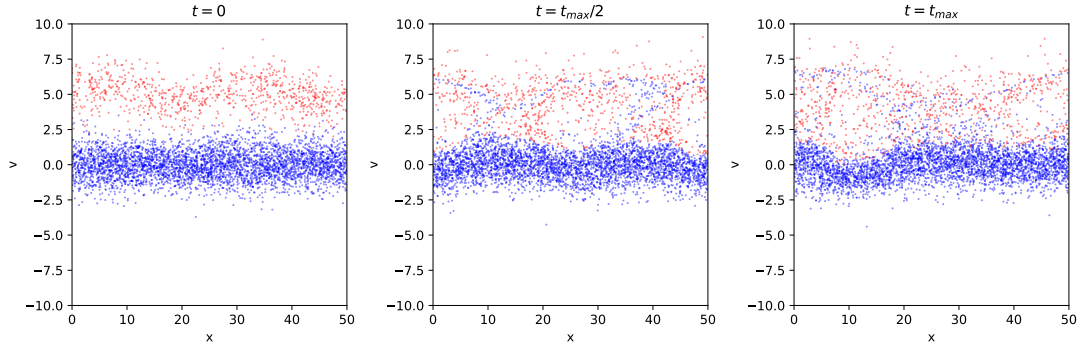

Fig. 2. The electron distribution in a phase space corresponding to Fig 1. Red dots denote high-energy electrons, and blue dots denote thermalized electrons.

Let $\mathbf{x} := [x_1; x_2; \cdots; x_n] \in \mathbb{R}^n$ and $\mathbf{v} := [v_1; v_2; \cdots; v_n] \in \mathbb{R}^n$. The electric potential $\Phi_h(\mathbf{x}) \in \mathbb{R}^{n_{mesh}}$ is given by Poisson equation with finite element matrix $\Lambda(\mathbf{x}) \in \mathbb{R}^{n \times n_{mesh}}$ while $\Lambda(\mathbf{x})^i_l := \lambda_i(x_l)$ and $l_n := [1; 1, \cdots; 1] \in \mathbb{R}^n$.

$$\Phi(\mathbf{x}) = L^{-1}\Lambda(\mathbf{x})^T l_n \tag{5}$$

where L corresponds to the Laplace operator with a finite element basis $\{\lambda_j(x)\}_{j=1}^{n_{mesh}}$ so that $L_{i,j} := (d_x\lambda_i, d_x\lambda_j)$. The corresponding ODE is given by 6.

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{v} \\ -\partial_{\mathbf{x}}\Lambda(\mathbf{x})L^{-1}\Lambda(\mathbf{x})^T l_n + E_{in}(\mathbf{x}) \end{bmatrix} \tag{6}$$

C. Bump-on-tail instability

Bump-on-tail instability [2] is a complex kinetic instability induced by the plasma-wave interaction that leads to energy transfer from high- to low-energy particles. Suppose the high-energy electrons are injected into the thermalized plasma. As Figure 1, the distribution of electrons in a velocity space has a *bump* on the tail of the distribution, which can be described as Equation 7.

$$f_0(v) = \frac{1}{(1+a)\sqrt{2\pi}} e^{-\frac{v^2}{2v_{th}^2}} + \frac{a}{(1+a)\sqrt{2\pi}} e^{-\frac{(v-v_b)^2}{2v_{th}^2}} \tag{7}$$

Here, v_{th} denotes the thermal velocity, v_b is the beam velocity, and a is a ratio of the high energy particles. For an initial wave perturbation, we apply the small sinusoidal wave with the amplitude A, which is given by

$$f(x, v, t = 0) = (1 + A \sin \frac{2\pi n_{mode}}{L} x) f_0(v) \tag{8}$$

Near $v \approx v_{ph}$, where $\partial_v f(v)|_{v=v_{ph}} > 0$, the inverse of Landau damping, which transfers the energy from particles to the wave, leads to a growth of the wave perturbation and is oscillated at the nonlinear stage [5]. As in Figure 1, the kinetic energy of the tail is then transferred by quasi-linear diffusion [5], which leads to the plateau at $t = 50$. As a consequence, the phase-mixing can be observed in a phase space as Figure 2. As Figure 3, it turns out that the change of the field energy corresponds to the time-evolution of the electron velocity distribution as Figure 1.

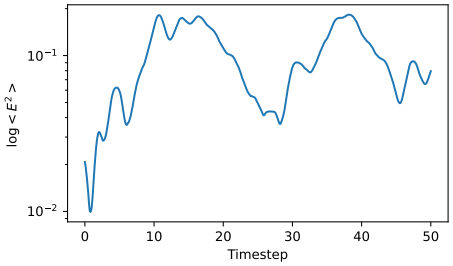

Fig. 3. The log-scale of the spatial-averaged electric field energy over time.

III. Optimal control problem

A. Cost functional for instability suppression

We formulate an optimal control problem to design an external electric field $E_{in}(\mathbf{x})$ that minimizes the growth of the instability. We define the cost functional J as

$$J := \int_{t_0}^{t_f} \frac{1}{2} \int_0^L |\nabla \phi|^2 dx dt + \frac{\lambda}{2} \int_{t_0}^{t_f} \int_0^L E_{in}^2 dx dt \tag{9}$$

The particle-wave interaction transfers energy from high-energy particles to the wave via the electric field, so that the electric field increases as the instability grows. Thus, the minimization of the electric field, which corresponds to first term in 9, should be considered to be reduced. The second term denotes the electric energy of the external field, which should be minimized to enhance overall efficiency while suppressing the instability.

We first parameterize the control input E_{in} as 2. Let $\mathbf{a} := (a_1, \cdots, a_n)^T$ and $\mathbf{b} := (b_1, \cdots, b_n)^T$. Then, due to orthogonality, the second term of 9 is given by $\frac{L}{4} \int_{t_0}^{t_f} |\mathbf{a}|^2 + |\mathbf{b}|^2 dt$. The first term can be discretized by following 5, which yields $\int_0^L |\nabla \phi|^2 dx dt = -l_n^T \Lambda(\mathbf{x}) L^{-1} \Lambda(\mathbf{x})^T l_n$.

$$\begin{aligned} J_h &= \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)\ dt \\ &:= \int_{t_0}^{t_f} -\frac{1}{2} l_n^T \Lambda(\mathbf{x}) L^{-1} \Lambda(\mathbf{x})^T l_n + \frac{\lambda}{4} |\mathbf{u}|^2\ dt \end{aligned} \tag{10}$$

Here, $\mathbf{u}(t) = \big[\mathbf{a}(t); \mathbf{b}(t)\big]$ is a control input parameter, and $L(\mathbf{x}, \mathbf{u}, t) = -\frac{1}{2} l_n^T \Lambda(\mathbf{x}) L^{-1} \Lambda(\mathbf{x})^T l_n + \frac{\lambda}{4} |\mathbf{u}|^2$ is a running cost.

However, plasma kinetic system consists of a large number of particles and is nonlinear. It is difficult to solve 10 due to high-dimensionality and nonlinearity. Thus, general approaches, such as direct or indirect methods, are not well-suited to this problem. To address this, we examine how to connect our problem to *Principle of Optimality* via value function, which is still intractable though. Then, we approximate the value function using approximate dynamic programming, known as reinforcement learning [6], to tackle this problem.

B. Dynamic Programming

Dynamic Programming (DP) is a framework for solving optimal control problems through backward-in-time evolution. The main principle is *Principle of Optimality*, which leads to the Hamilton-Jacobi-Bellman (HJB) equation. The HJB equation establishes a fundamental connection between reinforcement learning and optimal control theory, which explains how reinforcement learning can solve optimal control problems.

1) Value Function: We begin by defining the value function $V(t, x)$, which represents the minimum cost starting from state x at time t and following an optimal control until $t = t_f$.

$$V(t, x) := \inf_u \{ K(X_{t_f}^{(t,x)}) + \int_t^{t_f} L(X_s^{(t,x)}, u_s, s)\ ds \} \tag{11}$$

where $K(X_{t_f}^{(t,x)})$ is a terminal cost and $L(X_s^{(t,x)}, u_s, s)$ is a running cost. $X_s^{(t,x)}$ denotes the state trajectory from initial condition (t, x) under control u_t. In our problem, the value function is

$$V = \inf_{\mathbf{u}} \int_t^{t_f} -\frac{1}{2} l_n^T \Lambda(\mathbf{x}_s) L^{-1} \Lambda(\mathbf{x}_s)^T l_n + \frac{\lambda}{4} |\mathbf{u}_s|^2\ ds \tag{12}$$

2) Principle of Optimality and HJB equation: Principle of Optimality states that an optimal policy ensures that from any point along its trajectory, the subsequent decisions are optimal given the current state.

Theorem (Principle of Optimality)**.** *For every $(t, x) \in (t_0, t_f) \times \mathbb{R}^m$ and every $\Delta t \in (0, t_f - t_0]$, the value function $V(t, x)$ defined as Equation 11 satisfies that*

$$\begin{aligned} V(t, x) = \inf_{u \in \mathbf{u}[t, t+\Delta t]} \{ \int_t^{t+\Delta t} & L(X_s^{(t,x)}, u_s, s)\ ds \\ & + V(t + \Delta t, X_{t+\Delta t}^{(t,x)})\ \} \end{aligned} \tag{13}$$

This recursive structure then yields the Hamilton-Jacobi-Bellman equation 14. This provides necessary and sufficient conditions for the optimality of a control.

$$-\partial_t V(t, x) = \inf_u \{ L(x, u, t) + \langle \nabla_x V(t, x), f(x, u, t) \rangle \} \tag{14}$$

where $V(t_f, x) = K(x)$ for $\forall x \in \mathbb{R}^m$. The optimal solution $V(t, x)$ of 13 is a classical solution of the HJB equation.

C. Reinforcement Learning

HJB equation is solved backwards in time, but in general, it does not have a classical (smooth) solution. Instead, approximated dynamic programming combined with neural networks, so-called *Reinforcement Learning* (RL), has been suggested to approximate the value function by data-driven way. Here, a concept called *return* is used instead of cost, so the objective is to maximize a cumulative return. We define a policy $\pi(x) : \mathbb{R}^n \to \mathbb{R}^m$, which maps the state and input control, and a trajectory $\tau := (x_0, u_0, \cdots)$, which is a sequence of states and inputs generated by π. Then, a cumulative return with a discount factor γ is given by

$$R(\tau) := \sum_{t=0}^{T} \gamma^t r_t = \sum_{t=0}^{T} \gamma^t R(x_t, u_t) \tag{15}$$

Again, the goal is to find the optimal policy that maximizes the expected return $J(\pi)$.

$$J(\pi) := \mathbb{E}_{\tau \sim \pi}[R(\tau)] \tag{16}$$

To do this, we need to define value function $V^\pi(x)$ and action-value function $Q^\pi(x, u)$, which are the expected return with respect to state x and a pair of state and input (x_t, u_t).

$$\begin{aligned} V^\pi(x) &:= \mathbb{E}_{\tau \sim \pi}[R(\tau) | x_0 = x] \\ Q^\pi(x, u) &:= \mathbb{E}_{\tau \sim \pi}[R(\tau) | x_0 = x, u_0 = u] \end{aligned} \tag{17}$$

Then, the optimal policy π^* follows

$$\begin{aligned} \pi^*(x) &= \arg\max_{u\sim\pi(\cdot)} Q^{\pi}(x,u) \\ V^{\pi^*}(x) &= \max_{\pi} \mathbb{E}_{\tau\sim\pi}[R(\tau)|x_0 = x] \\ Q^{\pi^*}(x,u) &= \max_{\pi} \mathbb{E}_{\tau\sim\pi}[R(\tau)|x_0 = x, u_0 = u] \end{aligned} \tag{18}$$

Note that $V^{\pi}(x)$ and $Q^{\pi}(x,u)$ are connected by 17 and 18.

$$\begin{aligned} V^{\pi}(x) &= \mathbb{E}_{u\sim\pi}[Q^{\pi}(x,u)] \\ V^{\pi^*}(x) &= \max_{u}[Q^{\pi^*}(x,u)] \end{aligned} \tag{19}$$

The optimal input $u_t^* = \arg\max_u Q^{\pi^*}(x,u)$ also maximizes V^{π}. It turns out that this V^{π} is the same as the value function in HJB equation. Because those functions satisfy *Bellman equation*, as 20, which has the same form as the HJB equation. Thus, the optimal V^{π^*} in RL can be a candidate solution to HJB equation.

$$\begin{aligned} V^{\pi^*}(x) &= \max_{u} \mathbb{E}_{x'}[r_t(x,u) + \gamma V^{\pi^*}(x')] \\ Q^{\pi^*}(x,u) &= \mathbb{E}_{x'}[r_t(x,u) + \gamma \max_{u'} Q^{\pi^*}(x',u')] \end{aligned} \tag{20}$$

By defining $J = \int_t^{t_f} \frac{1}{2} l_n^T \Lambda(\mathbf{x}_s) L^{-1} \Lambda(\mathbf{x}_s)^T l_n - \frac{\lambda}{4}|\mathbf{u}_s|^2 \, ds$ as RL objective, the HJB equation turns into a problem for maximizing the expected return for RL.

RL algorithms aim to solve Bellman equations when the dynamics and reward structure are unknown. They learn the value function and its corresponding policy through interaction with the environment. RL employs two components:

- Actor network: Learns π^* that achieves the optimal in the HJB equation, using policy gradients $\nabla_u Q(x,u)$ for improvement.
- Critic network: Approximates $V(t,x)$ from the HJB equation.

For continuous state and control spaces, Deep Deterministic Policy Gradient (DDPG) is used [7], an actor-critic, model-free algorithm based on the deterministic policy gradient for continuous domains.

IV. Numerical experiments

This section presents the numerical results of the PIC simulation for instability suppression with DDPG algorithm. We integrated PIC code with the DDPG algorithm as they interact with the simulation environment to get the plasma state $(\mathbf{x}, \mathbf{v}) \in \mathbb{R}^{2n}$ and compute the optimal input parameters $u = (\mathbf{a}, \mathbf{b})$ following 3 at each step. Figure 4 depicts the diagram of how DDPG learns optimal control.

We define our own return function $r(x_s, u_s)$, which corresponds to our problem, designed to minimize the cost function 9 by maximizing the expected return in RL framework.

$$\begin{aligned} r(x_s,u_s) &= \bar{r}_{ee}(x_s) + \lambda \bar{r}_{ie}(u_s) \\ &= (1 - \frac{r_{ee}(x_s)}{r_{ee}(x_0)}) + \lambda(1 - \frac{r_{ie}(u_s)}{r_{ie}(u_0)}) \\ r_{ee}(x_s) &= \frac{1}{2} l_n^T \Lambda(\mathbf{x}_s) L^{-1} \Lambda(\mathbf{x}_s)^T l_n \\ r_{ie}(u_s) &= -\frac{1}{4} \mathbf{u}_s^T \mathbf{u}_s \end{aligned} \tag{21}$$

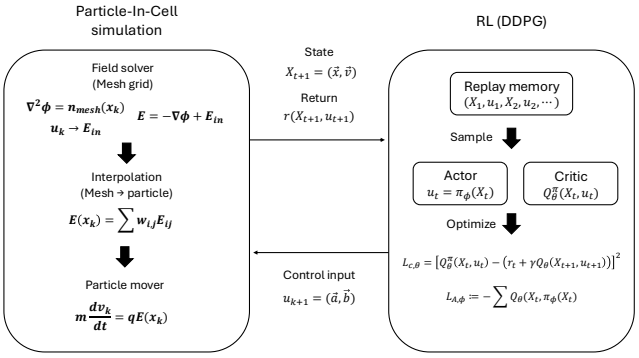

Fig. 4. Diagram for simplified PIC - DDPG optimization process

And Table I shows the parameters used in this experiment.

TABLE I
Parameters for simulation

Parameter	Value
Number of particles	$n = 10000$
Domain length	$L = 50.0$
Thermal velocity	$v_{th} = 1.0$
Beam velocity	$v_b = 5.0$
Wave amplitude	$A = 0.1$
High energy fraction	$a = 0.2$
Mode number	$n_{mode} = 2$
Input wave mode	$m = 3$
Time step	$\Delta t = 0.01$
Terminal time	$t_f = 50.0$
Cost coefficient	$\lambda = 1.0$
Learning rate	$lr = 0.0001$
Episode for training	$n_{episode} = 10000$
Discount factor	$\gamma = 0.995$

We now compare the uncontrolled and controlled cases through Fourier analysis of the system's electric field when a perturbed wave with $n_{mode} = 2$ is initially applied. Figure 7 shows the Fourier spectrum evolution without control input, while Figure 8 presents the corresponding spectrum with optimal control.

In the uncontrolled case (Figure 7), the $n = 2$ mode exhibits exponential growth during the linear phase ($t < 20$), characteristic of bump-on-tail instability driven by inverse Landau damping. The spectral energy accumulates at this resonant mode as energy is transferred from the high-energy beam particles to the wave. Subsequently, nonlinear effects lead to quasi-linear plateau formation in the velocity distribution, accompanied by saturation and oscillation of the mode amplitudes.

With optimal control (Figure 8), the $n = 2$ mode is significantly suppressed throughout the simulation. The optimal control strategy introduces a temporary enhancement of the $n = 1$ mode during the early phase ($t < 20$), which then decays with bouncing. This behavior suggests that the controller strategically redistributes spectral energy to stable modes that naturally damp, thereby preventing resonant amplification of

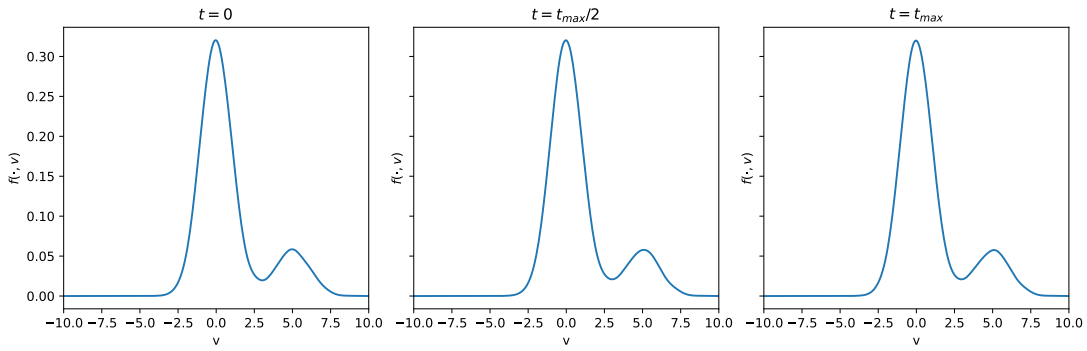

Fig. 5. The electron velocity distribution by the perturbed wave with $A = 0.2$, $n_{mode} = 2$, $v_{th} = 1.0$, and $v_b = 5.0$.

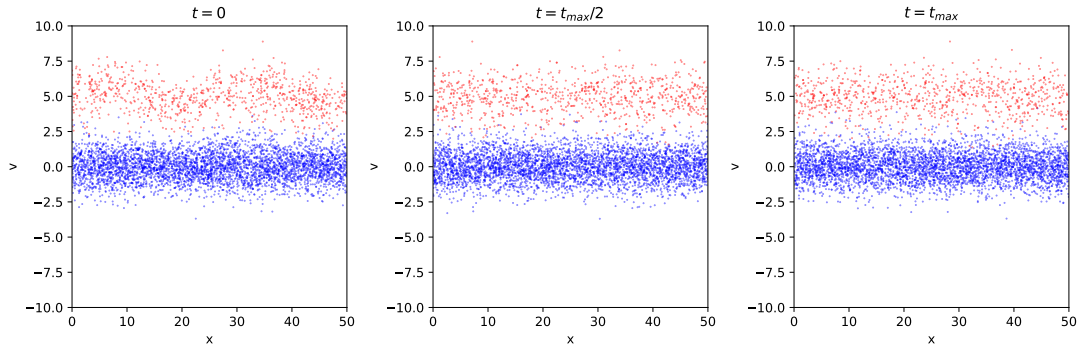

Fig. 6. The electron distribution in a phase space corresponding to Fig 1. Red dots denote high-energy electrers, and blue dots denote thermalized electrons.

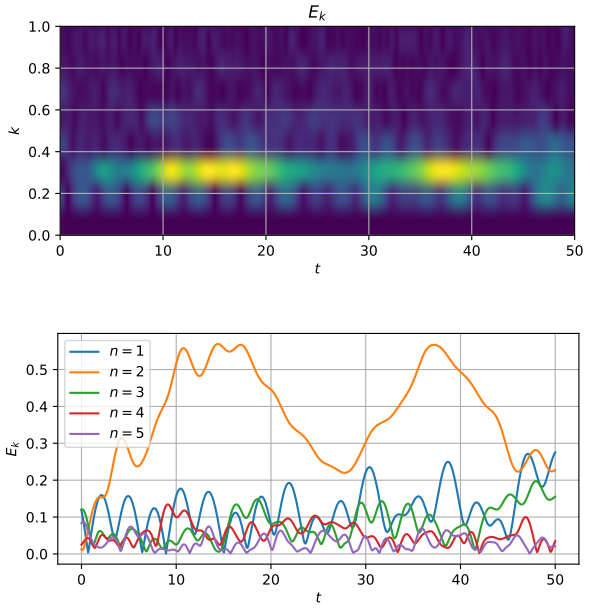

Fig. 7. Fourier spectrum of the electric field (upper) and Fourier coefficients corresponding to $n = 1$ to $n = 5$ without optimal control.

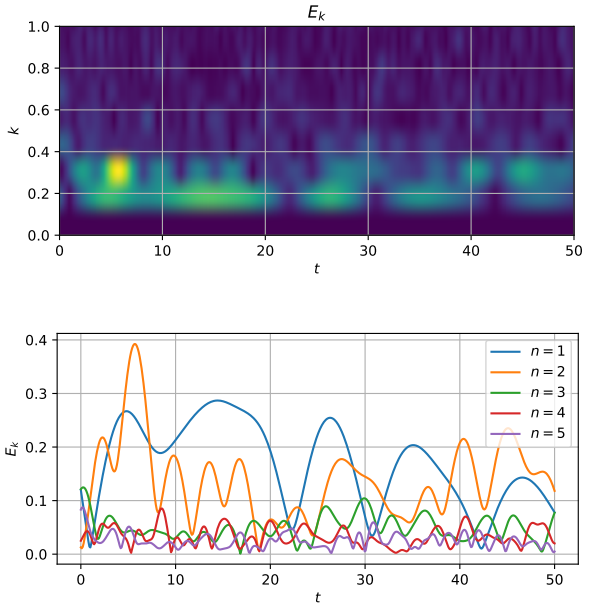

Fig. 8. Fourier spectrum of the electric field (upper) and Fourier coefficients corresponding to $n = 1$ to $n = 5$ with optimal control.

the unstable $n = 2$ mode. The controlled electric field (Figure 9) shows that the external input predominantly activates the $n = 1$ and $n = 2$ modes with time-varying amplitudes, adapting dynamically to counteract the instability growth.

TABLE II
COMPARISON OF ESTIMATED LANDAU DAMPING RATE

Parameter	Value
Without control	$\gamma_L = 0.00557$
Optimal control	$\gamma_L = 0.00034$

We then computed the linear Landau damping rate γ_L. Since the electric field energy $\log\langle E^2\rangle$ is proportional to $e^{-2\gamma_L t}$ at the linear stage, we can estimate the damping rate based on linear regression. Table II shows the estimated values for both uncontrolled and controlled cases. We can see that the optimal control with DDPG successfully suppresses the instability.

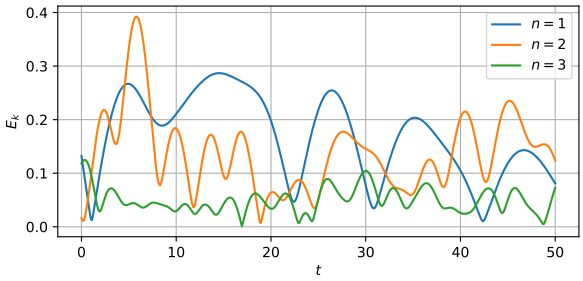

Fig. 9. Fourier coefficients $(a_i^2 + b_i^2)$ of the external electric field (control input) with $m = 3$ over time.

The effectiveness of instability suppression is evident in the velocity distribution and phase space evolution. Figure 5 demonstrates that the optimal control prevents the quasi-linear diffusion that would otherwise flatten the velocity distribution bump, maintaining the initial beam structure. Correspondingly, Figure 6 shows that the fine-scale phase-mixing observed in the uncontrolled case is largely prevented. This indicates that the optimal control interrupts the wave-particle resonance mechanism before significant energy transfer can occur.

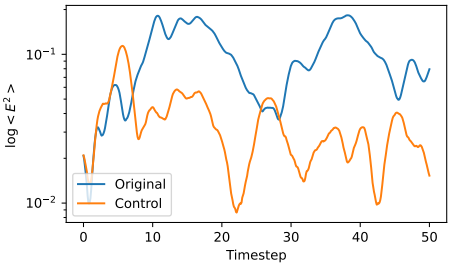

Fig. 10. Comparison of the spatial-averaged electric field energy over time between uncontrolled (blue) and controlled (orange) cases in logarithmic scale.

Figure 10 describes the comparison of the spatial-averaged electric field energy between the original and controlled cases. As mentioned in the previous section, the oscillation of the field energy is due to plasma-wave interaction through Landau damping. This means that the kinetic energy of high-energy electrons injected at $t = 0$ is transferred via Landau damping. However, the net field energy does not increase, as in the original case, indicating that the suppression of wave instability growth is effective.

V. CONCLUSION

This project demonstrates the successful application of reinforcement learning as a possible way to find the optimal control solution to suppress the bump-on-tail instability by applying external electric fields. We formulated the optimal control problem for instability suppression and connected the RL framework to the Bellman optimality equation corresponding to our problem. Our numerical experiments using the DDPG algorithm integrated with Particle-In-Cell simulations show that the learned control policy effectively suppresses the instability associated with the $n = 2$ mode, as validated by numerical analysis. We further extend this work to parameteric PIC simulations, in which different parameter configurations can also be covered.

REFERENCES

[1] Jeffrey P Freidberg. *ideal MHD*. Cambridge University Press, 2014.

[2] Luiz Fernando Ziebell, Rudi Gaelzer, and Peter H Yoon. Nonlinear development of weak beam–plasma instability. *Physics of Plasmas*, 8(9):3982–3995, 2001.

[3] Donald Gary Swanson. *Plasma kinetic theory*. Crc Press, 2008.

[4] Giovanni Lapenta. Particle in cell methods. In *With Application to Simulations in Space. Weather*. Citeseer, 2016.

[5] I. Dodin. Plasma waves and instabilities, 2025.

[6] Dimitri P Bertsekas. Neuro-dynamic programming. In *Encyclopedia of optimization*, pages 1–6. Springer, 2025.

[7] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.