Online reinforcement learning control of beam collision for BEPCII

Jiaqi Fan*, Jiuqing Wang**

Institute of High Energy Physics, CAS, 100049, Beijing, China

Abstract

For the Beijing Electron-Positron Collider II (BEPCII), operators need to tune the transverse offsets—including displacement and angular deviation (x, x', y, y')—of the two beams at the interaction point (IP) to maintain high luminosity as the beam current decays during normal operation. Given that the optimal offset exhibits a non-linear variation with beam current within a single run and also differs across individual runs, sustaining the optimal beam offset at the IP for consistent high luminosity at all times is laborious. Consequently, operators typically adopt a linear model for automatic offset tuning. In this study, a Deep-Q-Network (DQN) agent was trained using historical data to adjust the beam offset at the IP. The DQN agent employs 18 input parameters (including IP offset, beam position monitor (BPM) readings, and beam current) and 8 output parameters (Q-values for action selection). This DQN agent has been successfully deployed in daily offset tuning, essentially replacing both the linear model and manual operator adjustments. Furthermore, it has achieved an increase in integrated luminosity compared to the previous approach.

Introduction

BEPCII is a double ring collider working in 'decay' model

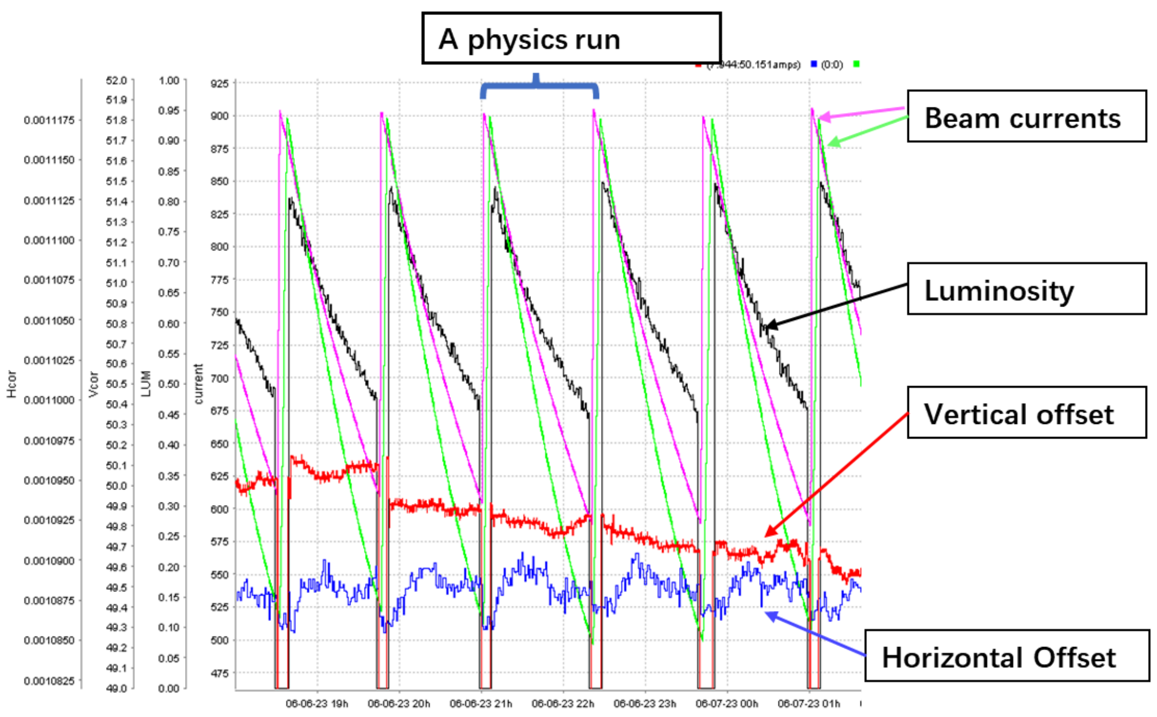


Fig 1: Schematic diagram of the offset adjustment at the IP of BEPCII

Offset Control Knobs

- Control the displacement and angular deviation at IP
- Four knobs (x,x',y,y') —— eight correctors (2 Bumps) for each ring
- The most frequently used luminosity optimization knobs
- Always only tune the knobs of electron ring and do not use x'
- Depends on orbit and current, need continuous optimization
- Tune manually ——scan one by one

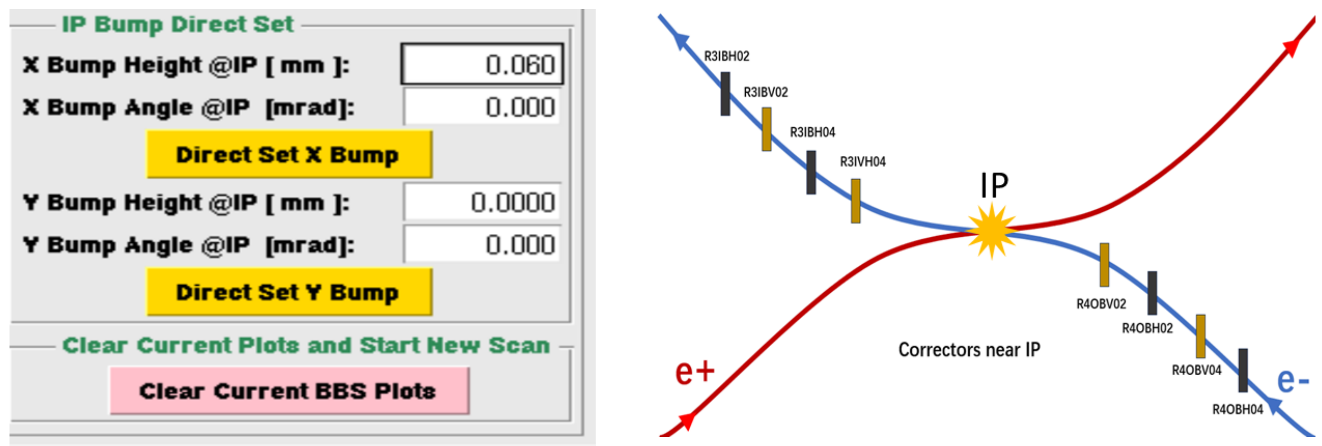


Fig 2: Offset knobs at the IP of BEPCII and the correction magnets used

RL method

Framework

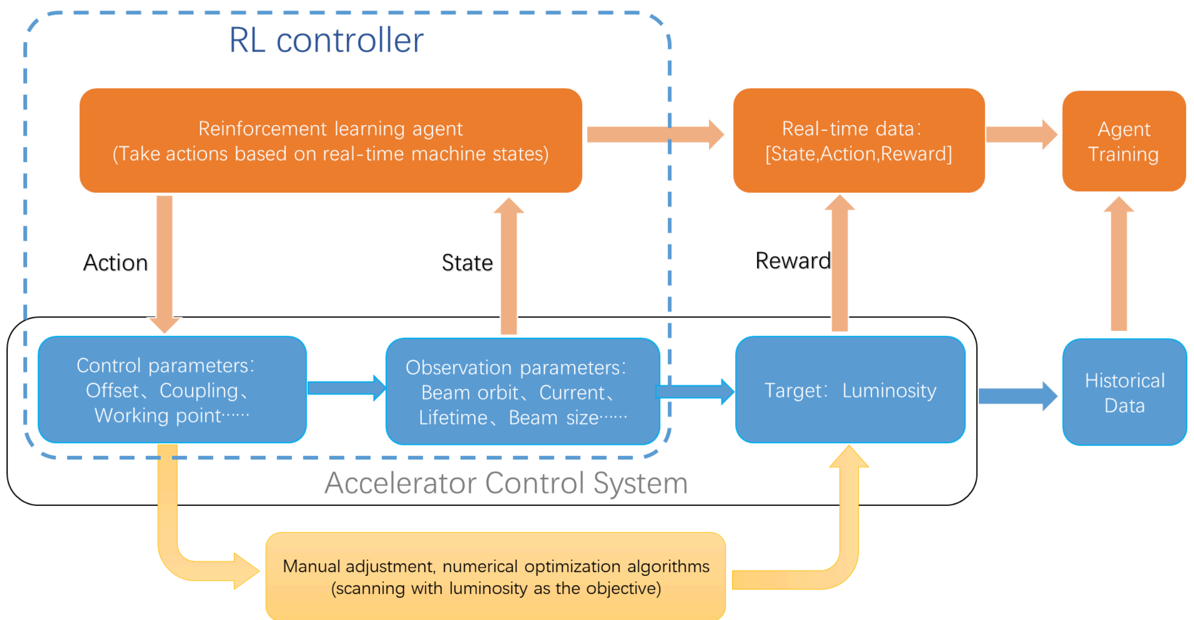


Fig 3: Main framework of the reinforcement learning controller

Why we use DQN?

- DQN is one of the most data-efficient algorithms
- Luminosity optimization simulation is difficult, making it hard to obtain effective information from the simulated environment
- The data interaction efficiency of BEPCII is low, resulting in limited available online training data
- The overall noise level is high, and excessively small adjustment steps are easily obscured by noise.

DQN Environment

State: [current , orbit , offset] — 18dims

Action: [x, y, y'] → [a1,a2,a3,a4,a5,a6]—6dims

Reward: Small-angle luminosity

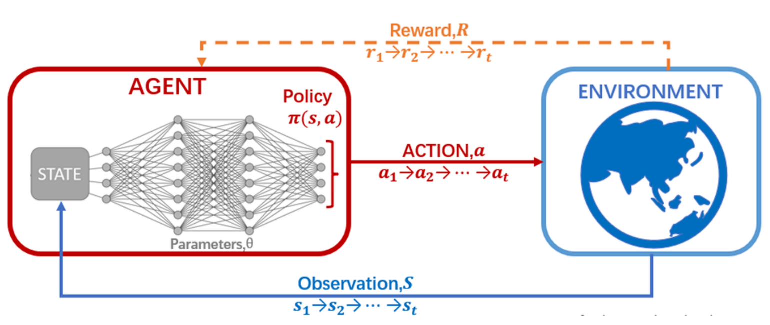
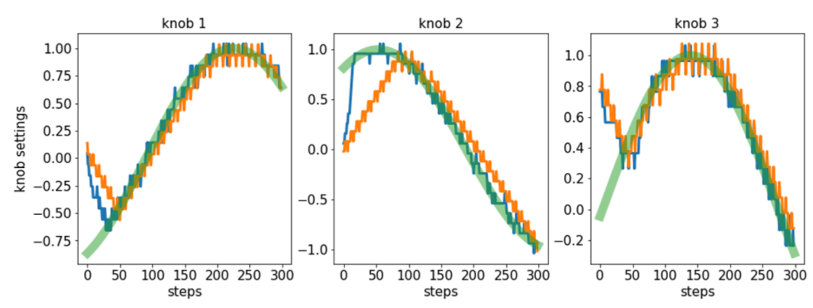


Fig 4: DQN framework

DQN agent training

Due to the lack of online data, we use a dither optimization algorithm to obtain a certain amount of training data for DQN pre-training, enabling it to complete online training faster and more safely.



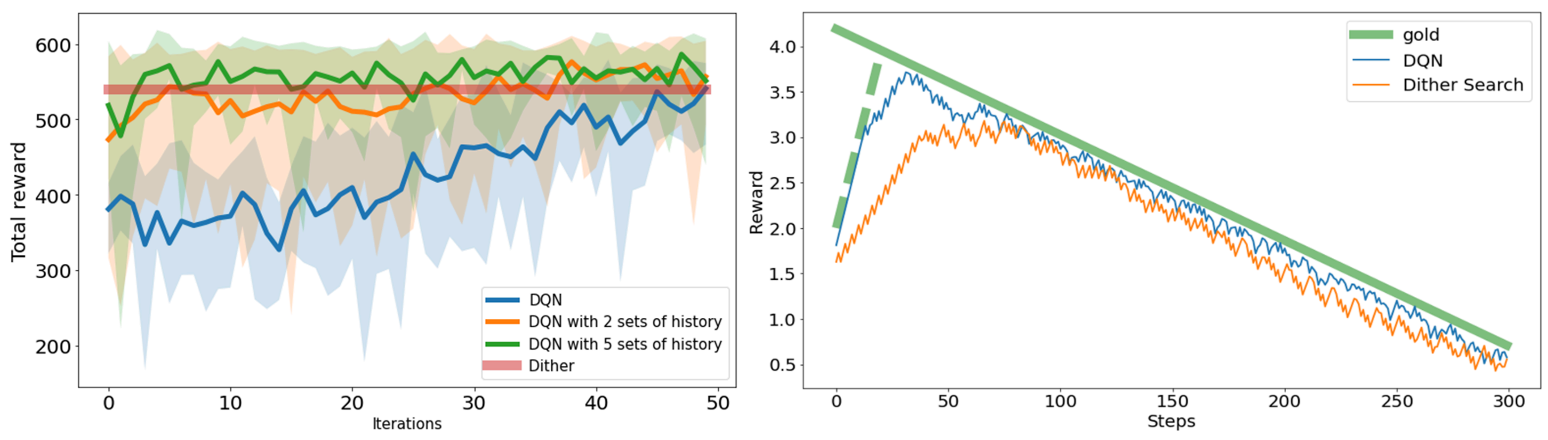


Fig 5: Simulation of DQN pre-training using historical data from dither optimization

Result

The DQN agent has been successfully used in the daily operation of BEPCII, almost replacing manual operations by operators and also bringing a certain degree of improvement in integrated luminosity.

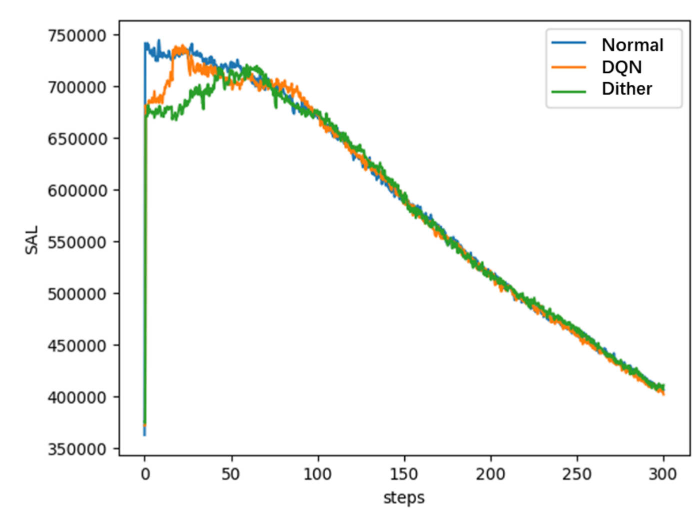


Fig 7: Comparison of luminosity between DQN and the dither optimization under large initial deviations

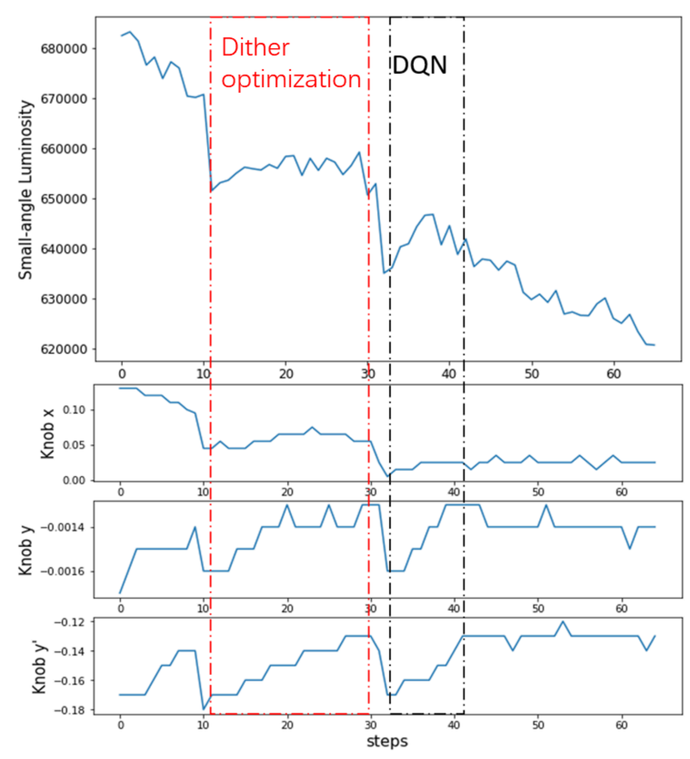


Fig 6: Comparison of optimization speed between DQN and the dither optimization

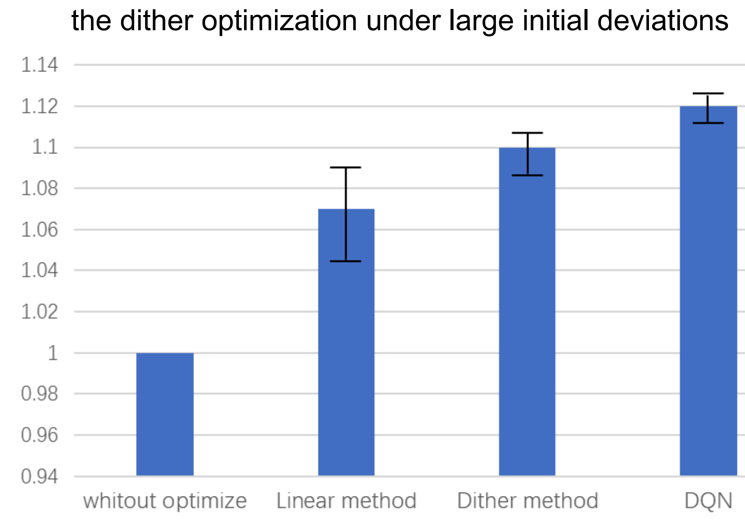


fig 8: The luminosity comparison under different control methods

*fanjiqi@ihep.ac.cn;**wangjq@ihep.ac.cn