ISBN: 978-3-95450-205-9 doi:10.18429/JACoW-Cyclotrons2019-TUP030

# REINFORCEMENT LEARNING BASED RF CONTROL SYSTEM FOR ACCELERATOR MASS SPECTROMETRY

H. Kim[1], M. Ghergherehchi[2], J. Lee[1], D.-H. Ha[2], H. Namgoong[2], K. M. M. Gad[1], J.-S.Chai[2,†]
Sungkyunkwan Universtity, 2066, Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do, Korea
[1]Dept. Energy Science
[2]Dept. Electrical and Computer Engineering

*Abstract*

Accelerator Mass Spectrometry (AMS) is a powerful method for separating rare isotopes and electrostatic type tandem accelerators have been widely used. At SungKyunKwan University, we are developing AMS that can be used in a small space with higher resolution based on cyclotron. In contrast to the cyclotron used in conventional PET or proton therapy, the cyclotron-based AMS is characterized by high turn number and low dee voltage for high resolution. It is designed to accelerate not only $^{14}C$ but also $^{13}C$ or $^{12}C$. The AMS cyclotron RF control model has nonlinear characteristics due to the variable beam loading effect of the acceleration of various particles and injected sample amounts. In this work, we proposed an AMS RF control system based on reinforcement learning. The proposed reinforcement learning finds the target control value in response to the environment through the learning process. We have designed a reinforcement learning based controller with RF system as an environment and verified the reinforcement learning based controller designed through the modelled cavity.

## INTRODUCTION

Accelerator mass spectrometry is an instrument for analysing the mass of radioactive isotopes. It accelerates various ions such as $^{10}Be$, $^{14}C$, $^{28}Al$, $^{38}Cl$, $^{41}Ca$, and $^{129}I$ and it is used in clinical experiments. AMS has higher resolution than conventional Mass Spectrometry. For general mass spectrometry, it has 10 - 12 parts per trillion (ppt) level sensitivity, but for accelerator mass spectrometer it has a high sensitivity of 10 - 15 ppt and tandem accelerators type AMS has been widely developed. However, cyclotron-based AMS is still under study because of its potential for miniaturization and efficiency compared to existing tandem accelerators [1, 2].

Unlike conventional cyclotron used for PET or proton therapy, AMS cyclotron has relatively low voltage and high rotation number for resolution. There is also the feature of accelerating various kinds of particles. This feature leads to non-linearity in control and interferes with performing precise beam and RF control. The external environment of the accelerator is continuously changed according to the type of particles and the quantity of the incident sample. In order to solve such a problem, it is inappropriate as a control system based on the existing linear section.

Reinforcement learning is one of the methods of machine learning such as Supervised Learning and unsupervised learning. It is a way to improve the behaviour through reward based on mutual relation of environment. Reinforcement learning does not require prior knowledge of the environment and is used for robots and games because it guarantees learning and adaptability [3].

In this work, reinforcement learning based RF control system was developed. From the viewpoint of reinforcement learning structure, we can redefine the controller as an agent and the cyclotron control variable as environment and to interconnect environment between agent, state, action and reward should be defined as shown in Fig. 1.

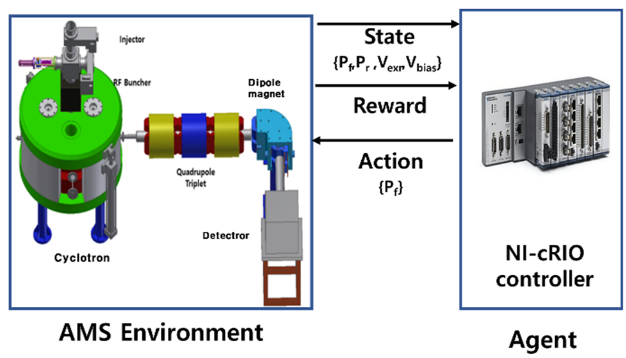



Figure 1: Reinforcement Learning based AMS control block diagram.

## SYSTEM DESIGN

In AMS RF cavity, the electric field from rf source can be calculated by following formula.

$$E_{PK} = \kappa_e\sqrt{P_t} \ ,$$

where $\kappa_e$ is coefficient which is determined using computer code and $P_t$ is transmitted power. $P_t$ is changed by cavity coupling coefficient and resonant-frequency mismatch and is related to beam loading effect and reflected power. Those parameters are used to describe state as follows:

$$\{P_f, P_r, V_{exr}, V_{bias}\} \quad ,$$

where $P_f$ is forward power from rf source, $P_r$ is reflected power and $V_{exr}$, $V_{bias}$ is extraction and biases voltage, which effect injection beam quality from ion source, respectively.

To measure forward and reflected power, ZFBDC20-61HP+ bi directional coupler was installed between cavity and RF amp and NI-cRio was communicated with ion source process controller station. The action space for the RF controller can be expressed by:

$$A = \{a| -\Delta f, 0, \Delta f\} \ ,$$

†jschai@skku.edu

ISBN: 978-3-95450-205-9 doi:10.18429/JACoW-Cyclotrons2019-TUP030



where $\Delta f$ is the Increment of RF input frequency. Since we use a relatively low frequency, we choose an AFG 3021 function generator as an RF input and ACTION is implemented by remote control. Agent's actions are evaluated by the reward:

$$\text{Reward} = \{+1, \text{ if } P_{r\,t+1} > P_{r\,t} \text{ and } -1, \text{ if } P_{r\,t+1} \leq P_{r\,t}\},$$

where $P_{r\,t+1}$, $P_{r\,t}$ are reflected power at two successive time steps. Agent will get positive reward when reflected power decrease. In other cases, it gets a negative reward.

## SOFTWARE DEVELOPMENT

AMS controller was implemented by NI CompactRIO systems. By using the Scan Engine, we programmed Actor-critic structure and data communication with other AMS devices for data acquisition and monitoring. CompactRIO consist of CPU module, input output module and communication module. The devices such as signal generator and ion source were connected through RS 232,485 serial communications.

In AMS controller, we programmed Actor-critic method [4]. Actor-Critic is type of reinforcement learning and also time difference method that can be used directly without an environment model, and has the advantage of doing the learning directly with raw experience.

In Actor-Critic structure, Both Actor and Critic are parameterized with neural networks and those function determines whether to pass through or update the neural network through Boolean input.

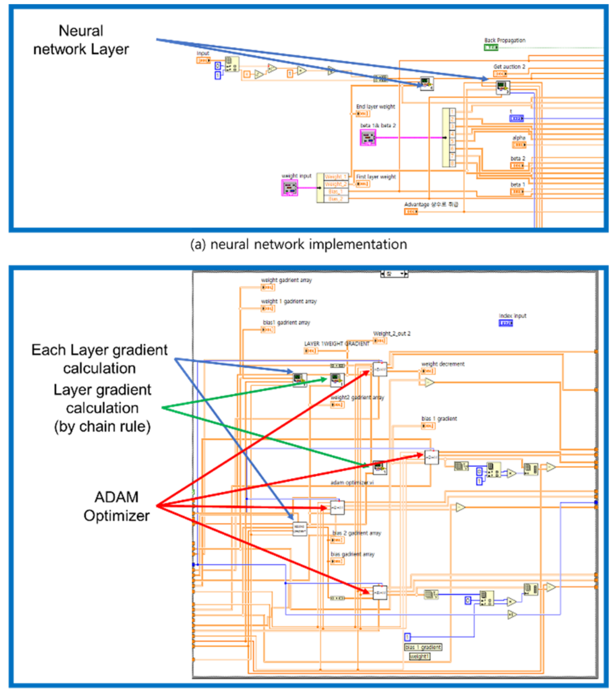
In the former case, the input data passes through each layer of the neural network. At this time, the weight and bias values remain at their stored values.

In the update process, the neural network calculates the gradient of the weight and bias according to the input through backpropagation, as shown in Fig. 2.

The calculated gradient is used as input to the ADAM optimizer [5] and updates the weights and biases. The weights and biases of neural networks that make up actors and critics are initialized before training. We performed the initialization using a He uniform variance scaling initializer. This is accomplished using a uniform white noise function with a square root of 6 / (number of layer inputs), as shown Fig. 3.

Actor and Critic perform optimization work by using ADAM Optimizer to reduce loss function through training. The ADAM optimizer is a kind of Stochastic Gradient Descent that allows you to quickly find the optimal value using only a few data when training. Figure 4 shows the loss function reduction during training of actors and critics.

Verification of the A2C-based resonant controller was carried out through the forward and reflected power. A2C controller was simulated when the resonance point did not change at 5.8 MHz to verify whether the A2C model can track the resonance point and confirmed that it converged to the 5.8 MHz band after the initial search, as shown Fig. 5.



(a) neural network implementation

(b) neural network weight and biases update process

Figure 2: Neural network block diagram.

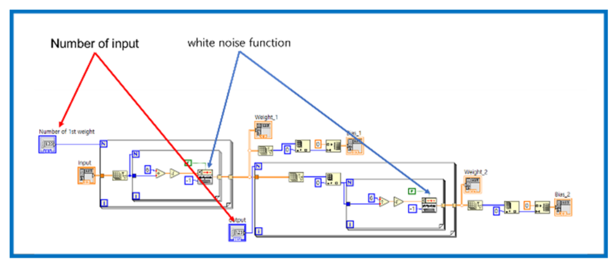
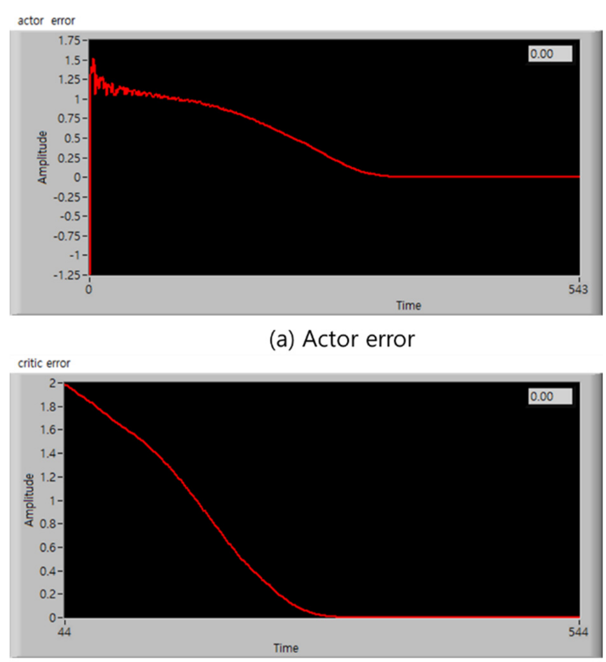


Figure 3: He uniform variance scaling initializer.



(a) Actor error

(b) Critic error

Figure 4: Actor and critic loss function test in training.

ISBN: 978-3-95450-205-9 doi:10.18429/JACoW-Cyclotrons2019-TUP030

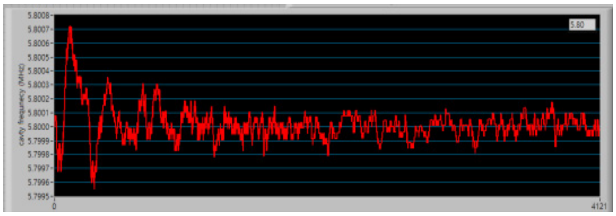


Figure 5: A2C Resonance controller learning process at 5.8 MHz constant resonance point.

The resonance point variation model was also trained by adding a sinusoidal function with an amplitude of 1.5 kHz to the resonance point, as shown Fig. 6.

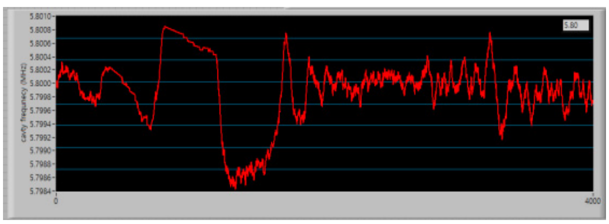


Figure 6: A2C Resonance controller learning process in resonance point shift model.

## DISCUSSION

A2C based AMS control system design and simulation was performed in this paper. This method is currently being tested with the ion source controller hardware information.

## ACKNOWLEDGEMENTS

This work was supported by the Radiation Technology R&D program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (Grant No. 2017M2A2A4A02020347).

## REFERENCE

[1] J.-W. Kim *et al.,* "Design study of a compact cyclotron for accelerator mass spectrometry", *J. Korean Phys. Soc.*, vol. 54, no. 2, pp. 580-585, Feb. 2009.

[2] W. Kutschera, "Progress in isotope analysis at ultra-trace level by AMS", *Int. J. Mass Spectrom.,* vol. 242, no. 2-3, pp. 145-160, Apr. 2005. doi.org/10.1016/j.ijms.2004.10.029

[3] R. S. Sutton and A. G. Barto, *"Reinforcement learning: An introduction"*, Cambridge, MA, USA and London, England: MIT press, 2014, 2015.

[4] I. Grondman *et al.*, "A survey of actor-critic reinforcement learning: Standard and natural policy gradients", *IEEE Trans. Syst. Man and Cybern. Part C Appl. Rev.,* vol. 42, no. 6, pp. 1291-1307, 2012. doi:10.1109/TSMCC.2012.2218595

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", 2014. arXiv:1412.6980