A Reinforcement Learning-based Online-training AI Controller for DC-DC Switching Converters

Xue Shi
School of Microelectronics
Northwestern Polytechnical University
Xi'an, China
shixue@mail.nwpu.edu.cn

Nan Chen
School of Electronics and Information
Northwestern Polytechnical University
Xi'an, China
nc@nwpu.edu.cn

Tingcun Wei
School of Microelectronics
Northwestern Polytechnical University
Xi'an, China
weitc@nwpu.edu.cn

Jiayu Wu
China Key System & Integrated Circuit Co., Ltd.
Wuxi, China
wujiayu_uestc@163.com

Peilei Xiao
China Key System & Integrated Circuit Co.,Ltd.
Wuxi, China
xiaoplcetc58@163.com

***Abstract*—A controller for DC-DC switching converters based on solely AI algorithm is proposed with a simpler structure than the traditional neural network-PID controllers. Reinforcement learning is used to train the AI controller online using deep deterministic policy gradient (DDPG) algorithm. The AI controller with an actor-critical architecture realizes model-free control with strong self-adaptive ability for different control objects, which can be used for different types of DC-DC switching converters. The performance of a buck DC-DC switching converter with the AI controller is compared with a neural network-PID controller through simulation. The simulation results show that the settling time is improved by at least 65% and overshoot/undershoot is decreased by at least 43%.**

Keywords—reinforcement learning, AI controller, actor-critical architecture, model-free, DC-DC switching converters.

I. INTRODUCTION

DC-DC switching converters are widely used in portable electronic devices, electric vehicles, and the Internet of Things applications (IoT) [1]-[4]. In recent years, in order to further improve the performance of digital power supplies and the adaptability of constant power loads (CPLs), intelligent digital controllers based on a neural network (NN) have attracted a lot of attention[5]. Compared with the traditional control methods, the NN-based controllers have the feature of faster response speed, wider adaptability and better robustness.

To train a NN-based controller for DC-DC switching converters, the deep learning is the most common way due to its simple structure [5]. However, it has to be trained through supervised learning, that is difficult to obtain the best calibration training data set. Therefore, in order to obtain the best control scheme under unsupervised learning, the reinforcement learning algorithm for training the NN was proposed to obtain relatively optimal controllers[6]-[10].

At present, the proposed reinforcement learning NNs for DC-DC converters are mainly divided into two categories. One is using the reinforcement learning NNs as an assist module to change the control parameters of traditional control schemes to improve the transient performance [6]-[9], however, the structure of this type controller is complex and its hardware cost is high. Besides, due to the traditional control schemes are linear control, compared with the nonlinear control, the transient performance is still limited. Another is using the reinforcement learning NN to generate the digital duty cycle directly [10]. In [10], an intelligent adaptive controller based on deep Q network (DQN) algorithm is proposed. Due to the characteristics of DQN, its performance is improved and its structure is simple, but it is unable to output the continuous duty cycle signal for DC-DC switching converters.

Aiming at improving the performance of model-free controller, this paper proposes a reinforcement learning-based online-training sole AI controller for DC-DC switching converters, in which the deep deterministic policy gradient (DDPG) algorithm [11] is adopted to train an Actor-Critic NN. Through setting the reasonable reward measures and the random conditions in the training process, the stability of the controller is improved. The remainder of this paper is organized as follows. Section II introduces the training environment of Actor-Critic NN controller based on reinforcement learning. Section III describes the Actor-Critic architecture and DDPG algorithm design. The simulation is presented in Section IV. Section V concludes the paper.

II. TRAINING ENVIRONMENT OF ACTOR-CRITIC NN CONTROLLER BASED ON REINFORCEMENT LEARNING

A. Buck DC-DC Basic Topology

The proposed controller only uses a neural network controller to regulate the DC-DC switching converters. In order to illustrate the proposed controller in more detail, we take the buck DC-DC switching converter as an example in this paper. Its basic topology is shown in Fig.1, V_g is the input voltage, S is the power switch, L_1 is the inductor, D_1 is the diode, C_1 is the capacitor, R_o is the output load, and $V_o(t)$ is the output voltage.

B. Overview of Reinforcement Learning Training Framework

The training environment for reinforcement learning NN-based controller is shown in Fig. 1. The ADC is used to convert

This work was supported in part by the National Natural Science Foundation of China under Grant 62101457 and in part by the Natural Science Foundation of Shaanxi Province, China under Grant 2021JM-064.

the analog output voltage $V_o(t)$ of the converter to the digital output voltage $V_o[k]$. The Actor-Critic NN-based controller generates the digital duty cycle $dnn[k]$. The DPWM converts the digital duty cycle $dnn[k]$ into an analog pulse signal $df(t)$. In addition to the buck DC-DC circuit in the training environment, the Actor-Critic NN-based controller to be trained also includes an observation module, a reward module, a stop module, and a circuit for generating reference voltage and errors to help the controller training.

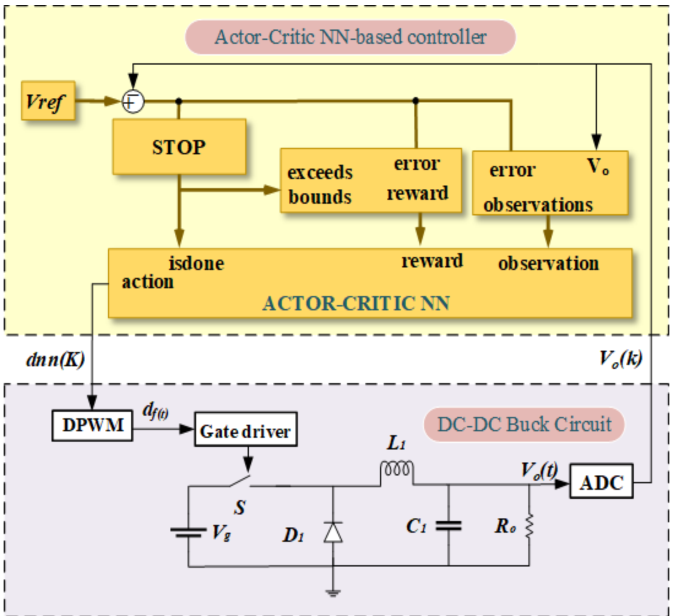

Fig. 1. Structure of training environment for reinforcement learning.

(1) *Observation Module:* For the Actor-Critic NN-based controller, at time t, the observation signal $S(t)$ is obtained through the observation module as the basis for output decision-making, so that the Actor-Critic NN-based controller can calculate the output duty cycle signal $dnn(k)$. In this paper, the observation module is designed to generate the observation vector $[\int edt, e, V_o]^T$, where V_o is the output voltage of the converter, $e=V_{ref}-V_o$ is the error between the output voltage and the reference voltage, $\int e\, dt$ is the integral of error e.

(2) *Reward Module:* Reinforcement learning calculates the immediate reward after the Actor-Critic NN controller outputs $dnn(k)$ through the reward function, which is the direct feedback of the interaction between the AI controller and the environment. The reward module constructs a scalar reward signal, in this paper, specifies the reward $r(t)$ as (1).

$$r(t) = 10(|e| < 0.01) - |e| - 100(e < -3.2||e > 1.8) \quad (1)$$

The reward signal is positive when the error is below 0.01 and negative otherwise. Also, there is a large reward penalty when the V_o is outside of a reasonable output for Buck DC-DC converter.

(3) *Stop Module:* Once the signal e exceeds a reasonable range in the training process, stop module will generate a stop signal to the NN controller to end this episode of training, and pass the stop signal to the reward module to generate a large reward penalty.

When the NN controller observes $S(t)$ and generates the action $a(t)$ which is actually $dnn(k)$ at time t, the reward module generates $r(t)$ to evaluate the control result at time $t+1$, so a sequence of vectors shown as (2) is obtained.

$$\{[S(0)\ a(0)\ r(0)], [S(1)\ a(1)\ r(1)], \\ \dots [S(n-1)\ a(n-1)\ r(n-1)], [S(n)\ a(n)\ r(n)]\} \quad (2)$$

It is worth notice that in the training process, in order to achieve better adaptability, V_{ref} and R_o will be reset randomly at the end of a training episode.

III. The Actor-critic Architecture Using DDPG Algorithm Design

A. Actor-Critic Architecture

The performance of the controller for DC-DC converter depends on a series of outputs *[dnn(0) ,…, dnn(n)]* of the controller, rather than one of them. It is necessary for AI controller to learn how to output a series of *dnn(k)* and obtain good performance. DDPG algorithm based on actor-critical architecture can solve this problem well.

The Policy network $\pi(a|s;\theta)$ is the Actor. It takes action a based on the state s, and the objective function $J(\theta)$ of policy learning is shown as (3), where τ is the sampling period, and $\pi_\theta(\tau)$ is the probability of sequence occurrence, and θ is parameter of the Policy network $\pi(a|s;\theta)$.

$$J(\theta) = E_{\tau\sim\pi_\theta(\tau)}[r(\tau)] = \int \pi_\theta(\tau) r(\tau) d\tau \quad (3)$$

Value network $Q(s,a;\theta)$ is the Critic. It scores the performance of Actor and quantifies the value of action a under the condition of state s. The relationship between Policy network (Actor) and Value network (Critic) is shown in Fig. 2.

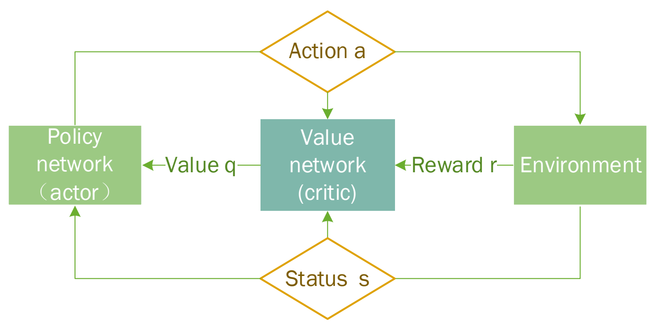

Fig. 2. Structure of actor-critic architecture.

Instead of the Actor, the reward r feedbacks to the Critic. The reason is that the objective function $J(\theta)$ of policy learning is the expectation of reward U, which is the weighted sum of all future rewards and not the expectation of reward r. Training the Policy network (Actor) requires the reward U, and the Value network (Critic) can estimate the expectation of reward U, so it can help to train the Policy network (Actor).

B. Taining Policy Network and Value Network by DDPG Algorithm

DDPG provides an excellent method for training the Police network and the Value network for Actor-Critic structure at the same time. Because the parameters of the Value network are often updated with gradients, they are also used to calculate the gradient of the Police network at the same time, which leads to instability in the learning process. Based on this, the DDPG algorithm creates two network copies for the Policy network and the Value network, one is online network and the other is target network.

After training a mini-batch of data, the parameters of the online network are updated through the SGA (Stochastic Gradient Descent with Averaging) or SGD (Stochastic Gradient Descent) algorithm, and then the parameters of the target network are updated through the soft update algorithm. The whole DDPG algorithm flow is shown in the Fig. 3.

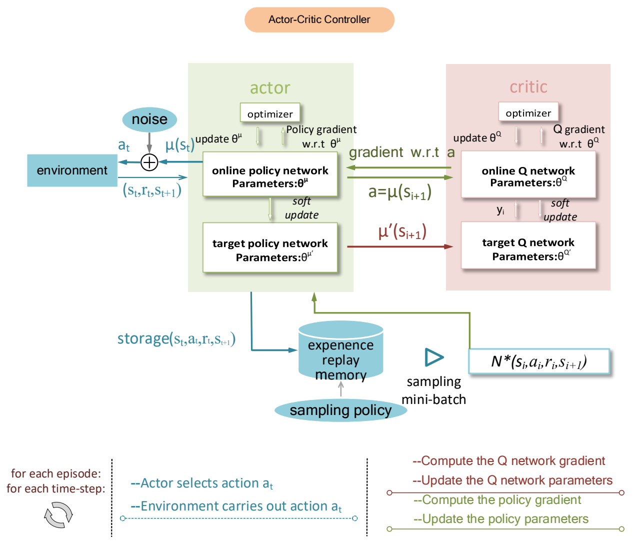

Fig. 3. The flow of DDPG algorithm.

IV. Verification and Analysis

To verify the performance of Actor-Critic NN-based controller, a buck DC-DC switching converter with the proposed controller is designed on the MATLAB/Simulink. Meanwhile, to illustrate the superiority of the proposed controller, the DC-DC switching converter using the neural network-PID controller is also designed for performance comparison. The parameters of the designed converter are given as follows: inductor $L_1=2.2\ \mu H$, capacitor $C_1=400\ \mu F$, switching frequency is 2.0 MHz, input voltage is 5.0 V, and output target voltage is 1.80 V.

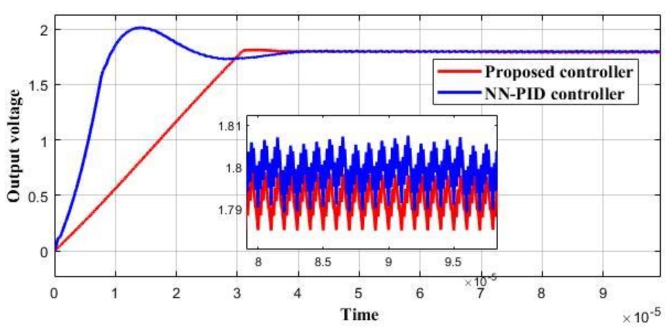

Fig. 4. The startup of the DC-DC switching converter using the proposed controller and the conventional NN-PID controller.

Fig. 4 shows the startup of the DC-DC switching converter using the proposed controller and the conventional NN-PID controller, respectively, the startup time and overshoot of the proposed controller are much better than that of the latter.

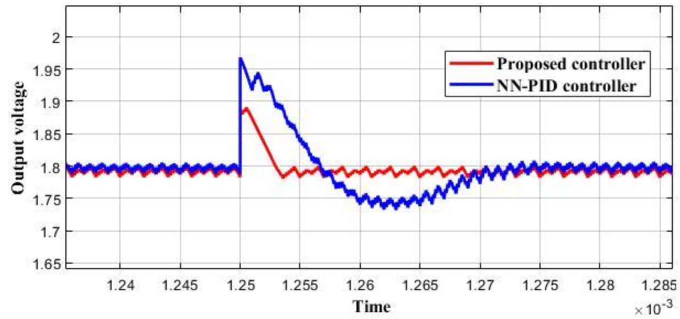

Fig. 5. The transient performance of DC-DC switching converters using the proposed controller and NN-PID controller when load current changes from 2.0A to 4.0A.

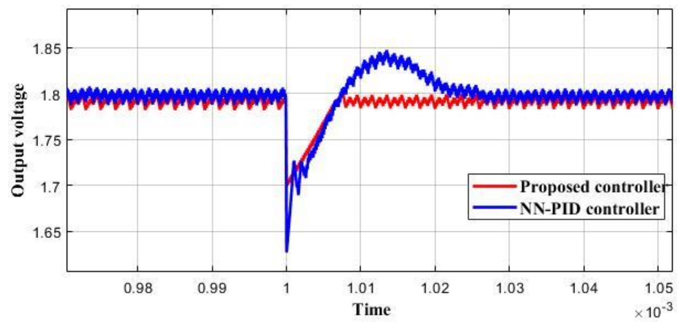

Fig. 6. The transient performance of DC-DC switching converters using the proposed controller and NN-PID controller when load current changes from 4.0A to 2.0A.

Fig. 5 and 6 shows the transient response performance of the DC-DC switching converter using the proposed controller and NN-PID controller. Obviously, the dynamic characteristics of the proposed controller are much better than that of the NN-PID controller. The detailed performance comparison between the proposed controller and the NN-PID controller is given in Table I.

TABLE I. Comparison of Transient Response Performances

	Performance index	*NN-PID*	*Proposed controller*	*Improved (%)*
	Startup(μs)	40	30	25.0
2.0→ 4.0A	Overshoot(mV)	175	100	42.8
	Settling time(μs)	20	7	65.0
4.0→ 2.0A	Undershoot(mV)	160	90	43.8
	Settling time(μs)	26	3	88.4

The proposed controller is also compared with other existing controllers. The detailed performance comparison is given in Table II. Due to the different switching frequencies, we use the regulation cycles to illustrate the settling time for the fairness of comparison. Apparently, the proposed controller has better transient characteristics.

V. Conclusion

In this work, a reinforcement learning-based online-training AI controller for DC-DC switching converters is presented to improve the dynamic characteristics. The proposed controller is a sole AI controller based on reinforcement learning that can output continuous duty cycle signal for DC-DC converters. The proposed controller adopts actor-critical architecture and is model-free control, which can be online-trained using DDPG algorithm. Moreover, this controller is with strong self-adaptive ability for different control objects, and can be used for many types of DC-DC switching converters. Simulation results show that the proposed controller has an excellent transient response performance. Compared with other controllers, the transient performances are significantly improved. In addition to some simulation results, the test on the hardware circuit for this controller will be carried out in the future.

TABLE II. PERFORMANCE COMPARISON

	Performance index	*[2]*	*[12]*	*[5]*	*[13]*	*[14]*	*This work*
	Output voltage (V)	5	1.8	20	14	24	1.8
	Switching frequency (MHz)	0.1	1	0.05	0.1	1	2
	Control method	NN-PID	FLC-PID	NN	SM	NN-MPC	Actor-Critic NN
	Inductor/Capacitor (μH/μF)	189/940	4.7/10	119/38	330/56	29/1.1	2.2/400
	Load current disturbance (mA)	800	200	333/157	200	2000	2000
Step-down	Overshoot (mV/mA)	0.100	0.188	0.525	0.455	2.400	0.045
	Settling time (cycle/mA)	0.150	0.060	0.060	0.012	0.008	0.007
Step-up	Undershoot (mV/mA)	0.112	0.212	1.592	0.500	2.700	0.050
	Settling time (cycle/mA)	0.175	0.115	0.159	0.013	0.008	0.003

REFERENCES

[1] N. Chen. “Digital Current-mode Controller Using Delta-Operator and Advance Sampling Predictive Control for High-Frequency DC-DC Switching Converters” IEEE J. of E. and S. Topics in Power Elec., Early access article, 2021, pp: 1-11.

[2] H. Maruta. “Characteristics Study of Neural Network Aided Digital Control for DC-DC Converter” The 2014 International Power Electronics Conference, pp: 3611-3615

[3] Kai-Yu Hu. “Digital V2 ConstantON-TimeControlBuckConverter With Adaptive Voltage Positioning and Automatic Calibration Mechanism” IEEE Trans. on Power Elec., Vol. 36, No. 6, Jun. 2021, pp: 7178-7188.

[4] M. Singh. “A 1-A 6-MHz Digitally Assisted Buck–Boost Converter With Seamless Mode Transitions and Fast Dynamic Performance for Mobile Devices” IEEE Trans. on Power Elec., Vol. 36, No. 4, Apr. 2021, pp: 4338-4351

[5] M. V. Martinovich. “Buck DC-DC Converter with Neural Network Sawtooth-Similar Carrier Signal Generator” 2018 19th Inter. Conf. on Micro/Nanotech. and Elec. Dev. Edm, pp: 629-634.

[6] M. Gheisarnejad. “A Novel Nonlinear Deep Reinforcement Learning Controller for DC–DC Power Buck Converters” IEEE Trans. on Ind. Elec., Vol. 68, No. 8, Aug. 2021, pp: 6849-6858.

[7] M. Hajihosseini. “DC/DC Power Converter Control-Based Deep Machine Learning Techniques: Real-Time Implementation” IEEE Trans. on Power Elec., Vol. 35, No. 10, Oct. 2020, pp: 9971-9978.

[8] Meysam G. “A Novel Deep Learning Controller for DC–DC Buck–Boost Converters in Wireless Power Transfer Feeding CPLs” IEEE Trans. on Ind. Elec., Vol. 68, No. 7, Jul. 2021, pp: 6379-6384.

[9] M Gheisarnejad. “IoT-Based DC/DC Deep Learning Power Converter Control: Real-Time Implementation” IEEE Trans. on Power Elec., Vol. 35, No. 12, Dec. 2020, pp: 13621-13630.

[10] C. Cui. “Voltage Regulation of DC-DC Buck Converters Feeding CPLs via Deep Reinforcement Learning” IEEE Trans. on Cir. and Sys. II: Express Briefs, Early access article, pp: 1-5.

[11] Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971 (2015).

[12] C Chang. “Field programmable gate array implementation of a single-input fuzzy proportional–integral–derivative controller for DC–DC buck converters” IET Power Elec., 2016, Vol. 9, Iss. 6, pp. 1259–1266

[13] H. Al-Baidhani. “Simplified Nonlinear Voltage-Mode Control of PWM DC-DC Buck Converter” IEEE Trans. on Energy Conv. Vol. 36, No. 1, Mar. 2021, pp: 431-440.

[14] J. Chen. “A Backpropagation Neural Network-Based Explicit Model Predictive Control for DC–DC Converters With High Switching Frequency” IEEE J. of E. and S. Topics in Power Elec., VOL. 8, NO. 3, Sep. 2020, pp: 2124-2142.