Residual DDPG Control with Error-Aware Reward Rescaling for Active Suspension Under Unseen Road Conditions

Zien Zhang[1], Abdul Hadi Abd Rahman[2]*, Noraishikin Zulkarnain[3]
Faculty of Information Science & Technology, UKM, Bangi, Selangor 43000 Malaysia[1]
Center for Artificial Intelligence Technology, UKM, Bangi, Selangor 43000 Malaysia[2]
Faculty of Engineering and Built Environment, UKM, Bangi, Selangor 43000 Malaysia[3]

***Abstract*—This study investigates a hybrid residual control framework combining Deep Deterministic Policy Gradient (DDPG) and a Proportional–Integral–Derivative (PID) based correction module for active suspension (AS) systems, aiming to improve ride performance and generalization under complex road excitations. The DDPG controller is trained on sinewave inputs, while the PID module compensates for residual errors to enhance robustness. To further guide policy optimization, an error-aware reward rescaling strategy is introduced during training, adaptively shaping the reward signal based on acceleration deviation. The controller is tested under five typical road conditions. These include sinewave inputs and step inputs, and ISO 8608 Level B random profiles. Simulation results show that the residual DDPG (RDDPG) controller works better than both DDPG alone and the PID controller. It reduces vertical acceleration RMS by 50.35% under a 0.05 m sinewave input. This shows that using reinforcement learning (RL) with fast correction and reward adjustment is a useful and stable way to control AS in different driving conditions.**

***Keywords*—Deep deterministic policy gradient; active suspension; reward function; generalization**

I. Introduction

In recent years, data-driven control methods such as RL have developed quickly. These methods have demonstrated strong adaptability and robustness in uncertain conditions [1]. At the same time, active anti-roll bar systems have shown strong potential in improving vehicle lateral stability [2]. Body roll can be reduced a lot during sharp turns or on rough roads [3], [4], [5]. These improvements show a new method to develop smart AS control. At the same time, more advanced methods are being tried in AS systems to better handle changing and unseen driving situations.

To enhance adaptability, Nguyen [6] developed a fuzzy logic PI controller that adjusts its gains in real-time, leading to improved ride comfort and stability in a quarter-car model under varying conditions. Similarly, Zhou et al. [7] introduced a machine learning model that incorporates fuzzy logic to mitigate vertical vibrations in a two degree of freedom (2-DOF) suspension setup. The method lowered body acceleration well and stayed simple to compute, so it could be used in real-time.

As the vehicle systems become more nonlinear and sophisticated, conventional controllers don't always guarantee good performance. For this reason, researchers look into neural networks (NN) and deep learning for better suspension force control. Bongain and Jamett [8] developed a hybrid fuzzy logic NN controller tailored to the nonlinear dynamics of electro-hydraulic actuators. They compared this development with the PID-ANN controller and showed that they achieved superior tracking performance in input variations. Konoiko et al. [9] introduced a feedforward deep NN trained using supervised learning. The data were acquired from the optimal PID controller by full state feedback. An automatic structure optimization procedure was performed to improve generalization.

AS control has improved a lot in recent years as more people use RL [10]. RL helps create control strategies by letting the system learn from the environment. This means it does not need a set system model. Liang and Wei [11] applied the DDPG algorithm to both active and semi-AS systems. The controller gives continuous force signals to the actuators and uses a reward function that punishes large vertical acceleration and too much suspension movement. Their results showed that the DDPG method worked well and made the ride more comfortable in different driving situations. Yong et al. [12] made a switching learning system based on the Soft Actor-Critic (SAC) framework. The system was able to recognize various types of road disturbances in real-time and adjust its control strategy accordingly. The method was first tested in simulation, then applied to real road situation. It showed that it could handle different driving conditions well.

Researchers have used actor-critic (AC) setups to better deal with uncertainty and changes in the system. Fares and Bani Younes [13] implemented a Time-Difference Advantage Actor Critic (TD-A2C) algorithm in a model-free setup. Their controller worked without knowing the environment before. It used an actor network to make actions, and a critic to check them. This setup helped keep a good balance between control force and ride comfort. Chen and Chang [14] applied a similar AC framework to control uncertain discrete-time systems. By using an NN to estimate the value function, their method kept the policy stable and made control more accurate. In parallel, Li and Kalabic [15] introduced a safety RL approach was introduced, integrating model free learning with a safety monitoring component. Penalty functions were used to reduce unsafe behavior during training which helped the controller explore effectively within safety limits.

*Corresponding Author.

Robustness and real-time adaptability remain key challenges in AS applications. To address this, Radac and Lala [16] formulated the suspension control task as a zero sum differential game between the controller and a worst case disturbance. Their game theoretic approach ensured robustness during learning by explicitly modeling adversarial conditions. Complementing this, Wang, Tian, and Zhang [17] developed a two layer control architecture based on a vehicle model. They combined incremental PID control with a deep RL adaptive (DRLA) strategy to enable online tuning of suspension forces in response to heave, pitch, and roll motions. Their method improved both ride stability and road handling. Wang, Zhuang, and Yin [18] applied DRL techniques to a quarter-car model under bump disturbances, confirming the controller's effectiveness in managing transient shocks.

The main contribution of this study is the use of error-aware reward rescaling method to make the training of the DDPG controller faster and more stable. The residual framework is combined with DDPG and PID to reduce vertical acceleration under different unseen road conditions.

II. METHODOLOGY

This study employs a classic quarter car model 2-DOF, representing the vertical motions of the sprung and unsprung masses. To guide the learning process, an Error-Aware Reward Rescaling mechanism is designed, where vertical acceleration minimization is set as the primary objective, and suspension deflection and control effort are treated as secondary considerations.

A. Quarter-Car Model and Parameters

The quarter-car model with 2-DOF represents the vertical motions of the sprung mass and the unsprung mass, respectively. The system dynamic equations and schematic diagram are given as follows, in Fig. 1:

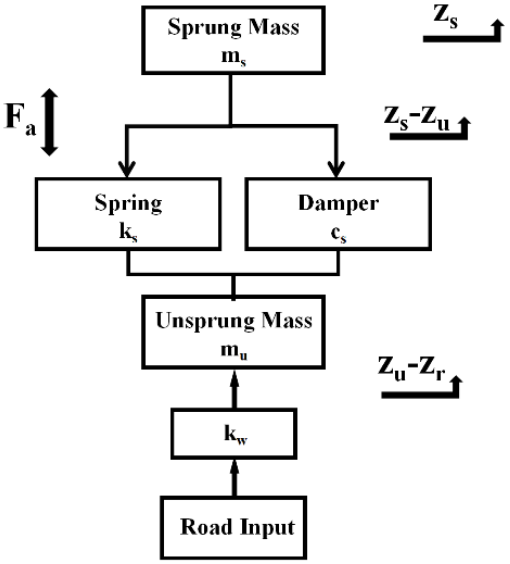

Fig. 1. Schematic diagram of the quarter-car AS system.

Sprung mass dynamics:

$$m_s\ddot{z}_s=-k_s(z_s-z_u)-c_s(\dot{z}_s-\dot{z}_u)+F_a \quad (1)$$

Unsprung mass (wheel) dynamics:

$$m_u\ddot{z}_u=-k_s(z_s-z_u)-c_s(\dot{z}_s-\dot{z}_u)-k_t(z_u-z_r)-F_a \quad (2)$$

where, m_s and m_w denote the sprung and unsprung masses, respectively; k_s and c_s represent the suspension stiffness and damping coefficient; k_w is the tire stiffness; z_s, z_u, and z_r denote the vertical displacements of the vehicle body, wheel, and road surface, respectively; and F_a is the control force generated by the AS actuator.

B. Control Framework and DDPG Controller

The control system developed in this study consists of a PID controller combined with an RL controller based on the DDPG algorithm. The DDPG controller generates the primary control force for the AS, while the PID controller learns to produce a residual correction term. The final suspension control force is the sum of the outputs from both controllers. This hybrid residual structure enhances the stability and responsiveness of the system during training as well as under unseen operating conditions. The overall control framework of the proposed AS system is illustrated in Fig. 2.

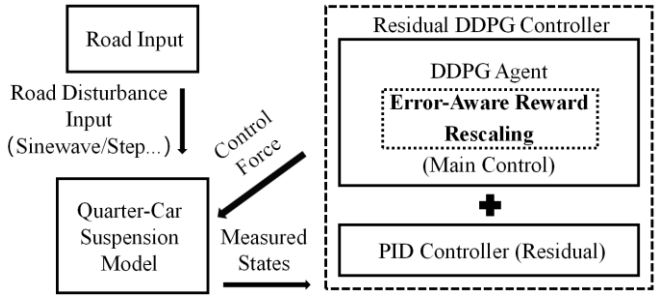

Fig. 2. Overall control framework of the AS system based on RDDPG.

The DDPG algorithm is a model-free RL method tailored for continuous control tasks. It integrates a deterministic policy gradient framework with an actor–critic architecture. The critic approximates the action-value function $Q(s, a;\theta^Q)$, which is updated by minimizing the temporal-difference loss between the estimated and target Q-values:

$$L(\theta^Q) = E[(y_t - Q(s_t, a_t; \theta^Q))^2] \quad (3)$$

The target value y_t is computed using delayed versions of the actor and critic networks:

$$y_t=r_t+\gamma\cdot Q'(s_{t+1},\mu'(s_{t+1})) \quad (4)$$

where, $\gamma\epsilon[0,1]$ is the discount factor, and μ', Q' are the target actor and critic networks respectively. These targets are held fixed for stability during mini-batch updates.

The actor represents the policy $\mu(s;\theta^\mu)$, which is updated by maximizing the critic's evaluation of the current policy. The gradient of the objective with respect to the actor's parameters is given by:

$$\nabla_{\theta^\mu}J\approx E\left[\nabla_a Q(s,a;\theta^Q)|_{a=\mu(s)}\cdot\nabla_{\theta^\mu}\mu(s)\right] \quad (5)$$

This update encourages the actor to select actions that yield higher predicted Q-values, as determined by the critic. To mitigate oscillations during learning, the algorithm employs

target networks that are updated gradually using a soft update rule:

$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}, \theta^{\mu'} \leftarrow \tau\theta^{Q\mu} + (1-\tau)\theta^{\mu'} \quad (6)$$

A small update coefficient τ ensures that target networks evolve slowly, thus smoothing the learning process. To promote exploration in a deterministic policy framework, temporally correlated noise is added to the actor's output:

$$a_t = \mu(s_t) + N_t \quad (7)$$

where, N_t typically follows an Ornstein–Uhlenbeck (OU) process, which models physical systems more realistically than uncorrelated noise and helps prevent premature convergence.

In this study, a residual structure is employed to improve the robustness of the DDPG controller. The RL agent is designed to fine-tune the output of the traditional controller rather than replace it entirely. The final control signal is obtained by combining the baseline controller's action with the correction learned by the agent, formulated as:

$$u_t = uPID(t) + uDDPG(t) \quad (8)$$

where, uPID denotes the output of a conventional PID controller, and uDDPG is the corrective term learned by the DDPG agent. This structure enables the RL component to focus on modeling deviations or nonlinearities that are difficult to capture with classical methods.

C. Error-Aware Reward Rescaling and Training Setup

This study introduces an Error-Aware Reward Rescaling mechanism that improves ride comfort by reducing vertical acceleration while accounting for suspension travel and control effort. The approach sets vertical acceleration as the primary objective and treats suspension deflection and actuator force as supporting factors. By adaptively shifting focus to the variable with the largest deviation at any given moment, it helps to guide the learning process more effectively, leading to better convergence and improved control reliability.

The instantaneous reward is composed of three terms: vertical comfort score $R_{comfort}$, suspension travel score R_{travel}, and control effort penalty P_{action}. The first two terms are modeled using exponential decay functions to penalize large deviations from ideal states:

$$R_{comfort} = \exp\left(-\left(\frac{|\ddot{z}_b|}{\ddot{z}_{lim}}\right)^2\right) \quad (9)$$

$$R_{travel} = \exp\left(-\left(\frac{|x_s|}{x_{lim}}\right)^2\right) \quad (10)$$

The control penalty is implemented using a hyperbolic tangent function to suppress excessively large control actions:

$$P_{action} = \tanh\left(\left(\frac{|F|}{F_{max}}\right)^2\right) \quad (11)$$

where, $\ddot{z}_b$ denotes the vertical acceleration of the vehicle body. x_s denotes the suspension travel. F denotes the actuator control force. $\ddot{z}_{lim}$ denotes the predefined upper limit for ideal vertical acceleration. x_{lim} is defined as the maximum desired suspension travel range. F_{max} denotes the maximum control force amplitude. λ refers to the control force penalty coefficient.

Instead of applying fixed weights to the comfort and travel terms, the proposed method adaptively rescales their importance based on the normalized error magnitude of each state component. Specifically, the dynamic weights are computed as:

$$w_c = \frac{\min\left(1, \frac{|\ddot{z}_b|}{\ddot{z}_{lim}}\right)}{\min\left(1, \frac{|\ddot{z}_b|}{\ddot{z}_{lim}}\right) + \min\left(1, \frac{|x_s|}{x_{lim}}\right)}, \; w_t = 1 - w_c \quad (12)$$

This formulation enables the controller to dynamically assign greater attention to the more critical performance objective at each moment, based purely on state error levels, without requiring access to prior reward outputs. As a result, the reward function guides the policy to minimize the dominant deviation and improve adaptability across varying conditions.

The final reward expression is given by:

$$r_t = w_c \cdot R_{comfort} + w_t \cdot R_{travel} - \lambda \cdot P_{action} \quad (13)$$

An additional bonus of 0.5 is granted when all normalized indicators fall below 0.5, encouraging convergence toward ideal operational zones. Numerical safety measures, including reward clipping and invalid value checks, are implemented to ensure stable learning. While the current reward function effectively guides the training process, the investigation into alternative reward shaping strategies lies outside the primary scope of this work and is reserved for future research.

The DDPG agent receives three state variables as input: suspension deflection, the corresponding vertical velocity of the sprung mass, and the vertical acceleration of the vehicle body. The output is a one-dimensional continuous control force ranging from −1000 N to 1000 N, which is applied to the AS system. Both the Actor and Critic networks adopt a single hidden layer architecture with 256 neurons. To enhance exploration, OU noise with a standard deviation of 0.5 is introduced. Detailed parameters are provided in the corresponding Table I.

TABLE I. DETAILED PARAMETERS OF DDPG CONTROLLER

DDPG Training Parameters	
Description	***Value***
Number of Hidden Neurons	256
Learning Rate (Actor & Critic)	0.001
Target Network Soft Update Factor	0.001
Exploration Noise Type	OU
Noise Standard Deviation	0.5
Experience Replay Buffer Size	1e6

The control input of the AS system is the suspension force generated by the RDDPG controller. This control structure consists of two components: a primary controller based on DDPG, which is trained under smooth road conditions, and a residual correction module based on a PID controller that compensates for the DDPG output. The controller receives system state variables as input, including the vertical acceleration of the vehicle body. The DDPG module outputs an initial control force, while the PID module learns the residual

component to refine the control output. This hybrid architecture enhances adaptability and robustness under previously unseen scenarios.

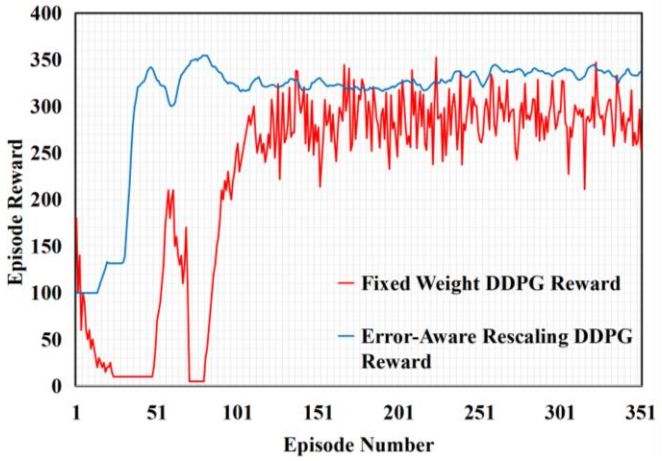

Fig. 3. Episode reward of Error-Aware Rescaling DDPG and Fixed Weight DDPG.

As shown in the training reward curve in Fig. 3, the DDPG agent utilizing the proposed Error-Aware Rescaling reward function demonstrates significantly superior learning performance compared to the baseline with fixed-weight rewards. Specifically, the rescaling-based agent achieves rapid convergence within the first 50 episodes and maintains consistently higher episode rewards with minimal fluctuations throughout the training process. In contrast, the fixed-weight reward agent exhibits unstable learning behavior, with prolonged periods of near-zero rewards and considerable oscillations even in later stages. This indicates that the fixed-weight design is prone to suboptimal policy learning and lacks robustness. Overall, the Error-Aware Rescaling approach enhances both convergence speed and policy stability, leading to improved final performance and demonstrating strong potential for guiding RL in suspension control tasks.

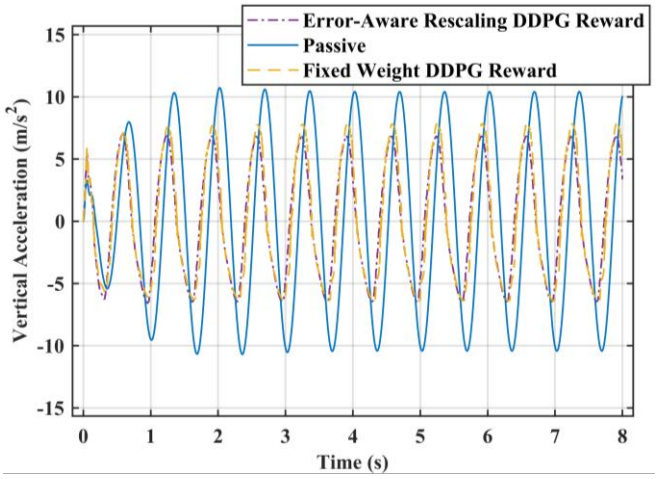

Fig. 4. Comparison of Error Aware Rescaling DDPG and Fixed Weight DDPG.

During training, a sinewave road excitation with a frequency of 1.5 Hz and amplitudes of 0.05m is used to simulate typical smooth driving conditions. The DDPG controller is trained until the policy converges, as indicated by a stable reward trend and a continuous reduction in the root mean square (RMS) of vertical acceleration.

TABLE II. SIMULATION RESULTS OF ERROR-AWARE RESCALING DDPG AND FIXED-WEIGHT DDPG

Controller	Compare DDPG at Sinewave 0.05m Road Input Condition	
	RMS	***RMS Reduction***
Passive	7.11 m/s²	-
Error-Aware Rescaling DDPG	4.62 m/s²	35.02%
Fixed-Weight DDPG	5.05 m/s²	28.94%

The results are shown in Fig. 4 and Table II. The passive suspension system produces an RMS value of 7.11 m/s², indicating significant vibration and poor ride comfort. When using the DDPG controller trained with the proposed error-aware rescaling reward, the RMS is reduced to 4.62 m/s², achieving a 35.02% improvement over the passive case. In comparison, the controller trained with fixed reward weights yields an RMS of 5.05 m/s², corresponding to a 28.94% reduction. The improved reward structure demonstrates a more pronounced effect in enhancing the vehicle's vertical stability and ride comfort.

III. SIMULATION SETUP

To evaluate the control performance and generalization capability of the proposed RDDPG control strategy in AS systems, this study conducts simulation research based on a quarter-car suspension model. The entire simulation environment is established within the MATLAB/Simulink platform. Both the training and testing procedures are carried out using a representative 2-DOF quarter AS model, which comprises the sprung mass, unsprung mass, suspension spring, damper, and active actuator. The system dynamics take into account the vertical motions of the vehicle body and wheel, aiming to characterize the suspension's response under various road excitations. Table III presents the detailed parameters of the vehicle.

TABLE III. DETAILED PARAMETERS OF VEHICLE

Vehicle Parameters	
Value	***Description***
1400 kg	Sprung mass (full car)
45 kg	Unsprung mass
25800 N/m	Suspension spring stiffness
1200Ns/m	Suspension damping coefficient
198000N/m	Tire stiffness

To evaluate the generalization capability of the proposed controller, three types of unseen road excitation profiles are introduced during the testing phase. The first type is a step road excitation, which includes step inputs of 0.03 m (unseen) and 0.05 m (trained) in amplitude. This test is designed to simulate sudden road height changes, such as speed bumps or curbs. The corresponding profile is illustrated in Fig. 5.

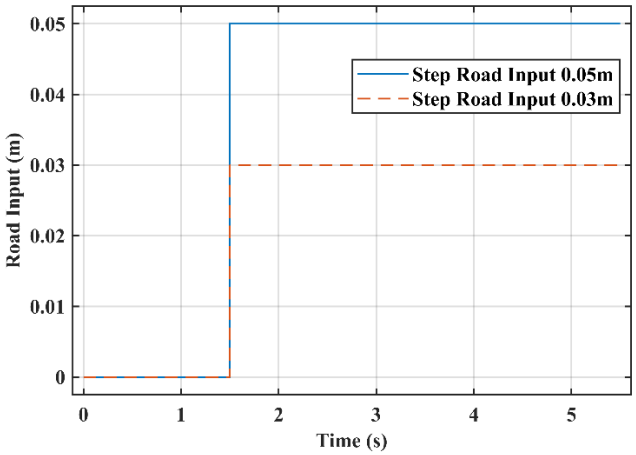

Fig. 5. Step 0.03m and step 0.05m road input profiles.

The second type is a random road excitation, constructed according to the ISO 8608 standard for road roughness classification. Specifically, Level B profiles are employed to represent medium and poor road conditions, respectively. These excitations are shown in Fig. 6.

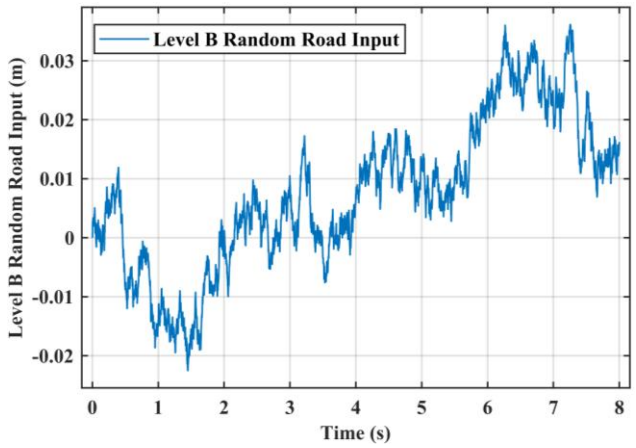

Fig. 6. Level B random road input profiles.

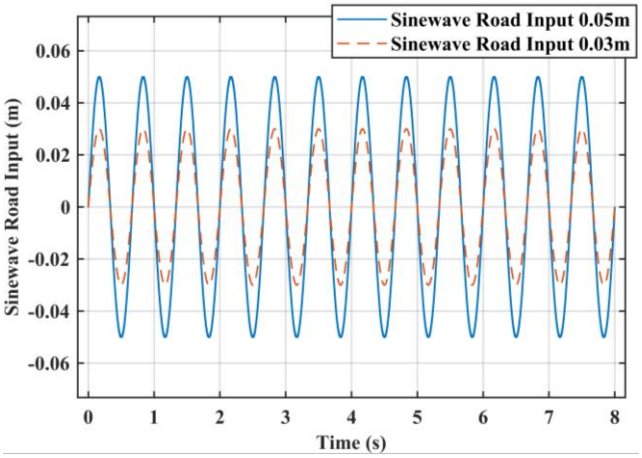

Fig. 7. Sinewave 0.05m and Sinewave 0.03m road input profiles.

The third type is a sinewave excitation, which is identical to the road profile used in the training phase. This sinewave 0.05m scenario serves as a baseline to validate the controller's fundamental performance under familiar conditions. This sinewave 0.03m is an unseen condition to validate the controller's performance. The excitation waveform is presented in Fig. 7.

IV. Simulation Results

To comprehensively assess the effectiveness and generalization ability of the proposed RDDPG control strategy under varying road conditions, three representative test scenarios were constructed: sinewave 0.05m input excitation (training condition), sinewave 0.03m input excitation (unseen condition), step 0.03m and 0.05m input (unseen condition), and ISO 8608 Level B random road profiles (unseen conditions). The figures present the vertical acceleration responses of the vehicle body under different control strategies, including passive suspension, PID control, standalone DDPG control, and the proposed RDDPG controller, when subjected to these road excitations.

A. Sinewave 0.05m Road Input Condition Simulation Results

The simulation results under the sinewave 0.05m road input condition are shown below:

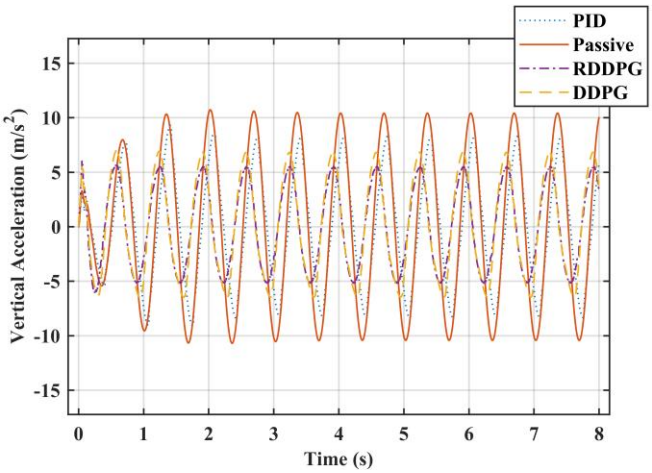

Fig. 8. Sinewave 0.05m road input simulation results.

During the training phase, a sinewave road excitation with an amplitude of 0.05 m was employed. Fig. 8 illustrates the vertical acceleration responses of the vehicle body under different control strategies. As shown, all three active control methods significantly reduce the fluctuation amplitude of vertical acceleration compared to the passive suspension, with the proposed RDDPG controller delivering the most pronounced improvement. According to the data summarized in Table IV, the passive suspension system RMS vertical acceleration of 7.11 m/s². The DDPG-based controller reduces this value to 4.62 m/s², corresponding to a 35.02% decrease, while the conventional PID controller achieves an RMS of 5.24 m/s², reflecting a 26.30% reduction. Notably, the RDDPG approach further lowers the RMS to 3.53 m/s², achieving a 50.35% reduction over the passive baseline and demonstrating the best overall performance in vibration attenuation. These results confirm the effectiveness and superiority of the RDDPG strategy under 0.05m sinewave excitation conditions.

TABLE IV. SIMULATION RESULTS OF SINEWAVE 0.05M ROAD INPUT CONDITION

Controller	Sinewave 0.05m Road Input Condition	
	RMS	***RMS Reduction***
Passive	7.11 m/s²	-
DDPG	4.62 m/s²	35.02%
RDDPG	3.53 m/s²	50.35%
PID	5.24 m/s²	26.30%

B. Sinewave 0.03m Road Input Condition Simulation Results

The simulation results under the sinewave 0.03m road input condition are shown below:

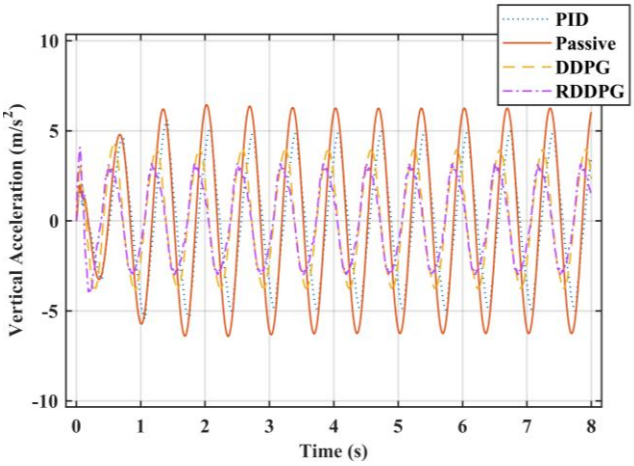

Fig. 9. Sinewave 0.03m road input simulation results.

To assess the vibration isolation performance under moderate road disturbances, a sinewave input with an amplitude of 0.03 m was applied in the simulation. The corresponding vertical acceleration responses for each control method are presented in Fig. 9. As observed, all AS strategies provide improved performance over the passive system, with the RDDPG controller demonstrating the most notable attenuation. According to the results summarized in Table V, the passive suspension yields an RMS vertical acceleration of 4.26 m/s². The DDPG controller reduces this to 2.84 m/s², indicating a 33.33% improvement, while the PID controller achieves 3.14 m/s², corresponding to a 26.29% reduction. The RDDPG strategy achieves the lowest RMS value of 2.14 m/s², marking a 49.77% decrease relative to the passive configuration. Overall, the results highlight the RDDPG controller's robust adaptability and its superior vibration suppression performance under reduced excitation amplitudes.

TABLE V. SIMULATION RESULTS OF SINEWAVE 0.03M ROAD INPUT CONDITION

Controller	Sinewave 0.03m Road Input Condition	
	RMS	***RMS Reduction***
Passive	4.26 m/s²	-
DDPG	2.84 m/s²	33.33%
RDDPG	2.14 m/s²	49.77%
PID	3.14 m/s²	26.29%

C. Step 0.05m Road Input Condition Simulation Results

The simulation results under the step 0.05m road input condition are shown below:

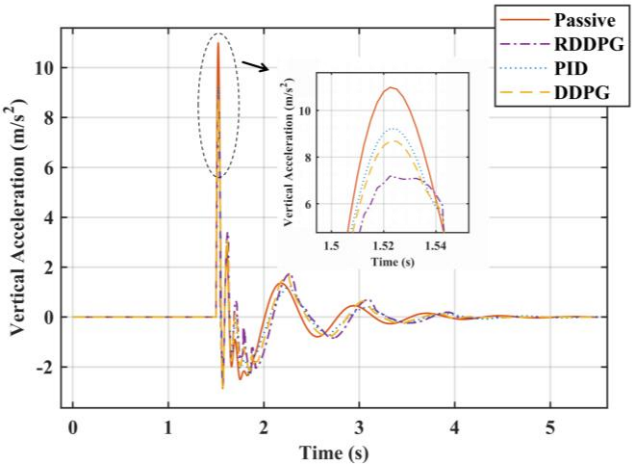

Fig. 10. Step 0.05m road input simulation results.

To evaluate the generalization performance of each control strategy under previously unseen conditions, a step road disturbance with a height of 0.05 m was introduced, simulating an abrupt change in road profile. The resulting vertical acceleration responses are illustrated in Fig. 10. Compared to the passive suspension, all three active control strategies clearly mitigate the initial impact and help the system stabilize more rapidly. The magnified section highlights the effectiveness of each controller in suppressing peak acceleration. The corresponding RMS values are listed in Table VI. The passive system yields an RMS vertical acceleration of 0.99 m/s². The DDPG controller reduces it to 0.82 m/s², resulting in a 17.17% decrease. The PID controller achieves 0.84 m/s², reducing the RMS by 15.15%. The RDDPG controller attains the best result with 0.79 m/s², representing a 20.20% reduction compared to the passive baseline. These findings demonstrate that the RDDPG strategy not only performs well under trained periodic conditions but also retains strong robustness and adaptability when facing sudden road disturbances.

TABLE VI. SIMULATION RESULTS OF STEP 0.05M ROAD INPUT CONDITION

Controller	Step 0.05m Road Input Condition	
	RMS	***RMS Reduction***
Passive	0.99 m/s²	-
DDPG	0.82 m/s²	17.17%
RDDPG	0.79 m/s²	20.20%
PID	0.84 m/s²	15.15%

D. Step 0.03m Road Input Condition Simulation Results

The simulation results under the step 0.03m road input condition shown below:

Fig. 11 illustrates the vertical acceleration responses of the vehicle body under different control strategies. Compared to the passive suspension, the active controllers effectively

suppress the initial impact peak and accelerate the system's convergence. The magnified section in the figure provides a detailed comparison of the overshoot behavior across controllers. Table VII summarizes the corresponding RMS values. The passive system yields an RMS vertical acceleration of 0.58 m/s². The DDPG controller reduces this to 0.48 m/s², representing a 15.79% improvement. The PID controller achieves 0.50 m/s², which corresponds to a 13.79% reduction. The RDDPG controller achieves the lowest RMS at 0.45 m/s², yielding a 21.05% decrease compared to the passive baseline. These results further demonstrate that the proposed RDDPG control strategy maintains reliable performance and strong generalization capability even under mild, abrupt disturbances.

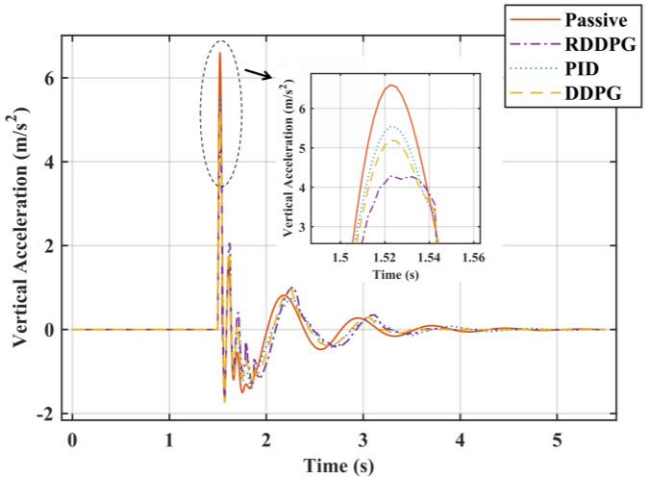

Fig. 11. Step 0.03m road input simulation results.

TABLE VII. SIMULATION RESULTS OF STEP 0.03M ROAD INPUT CONDITION

Controller	Step 0.03m Road Input Condition	
	RMS	***RMS Reduction***
Passive	0.58 m/s²	-
DDPG	0.48 m/s²	15.79%
RDDPG	0.45 m/s²	21.05%
PID	0.50 m/s²	13.79%

E. Level B Road Input Condition Simulation Results

The simulation results under the ISO 8608 Level B random road input condition are shown below:

The ISO 8608 Class B random road profile was introduced to examine how well each control strategy adapts to irregular and complex surface disturbances. Fig. 12 displays the resulting vertical acceleration responses. As shown, the passive suspension experiences pronounced high-frequency oscillations, whereas the active control strategies noticeably suppress these fluctuations, indicating their improved ride comfort capabilities. Table VIII reports the corresponding RMS values. The passive configuration results in an RMS of 0.76 m/s². The DDPG strategy improves this to 0.63 m/s², yielding a 17.11% reduction, while the PID controller achieves 0.66 m/s², reflecting a 13.16% decrease. RDDPG again provides the best result, further lowering the RMS to 0.62 m/s², an 18.42% improvement over the passive baseline. These findings suggest that the RDDPG approach retains a high level of performance even under stochastic mid-grade excitations, demonstrating greater resilience and adaptability than the other methods tested.

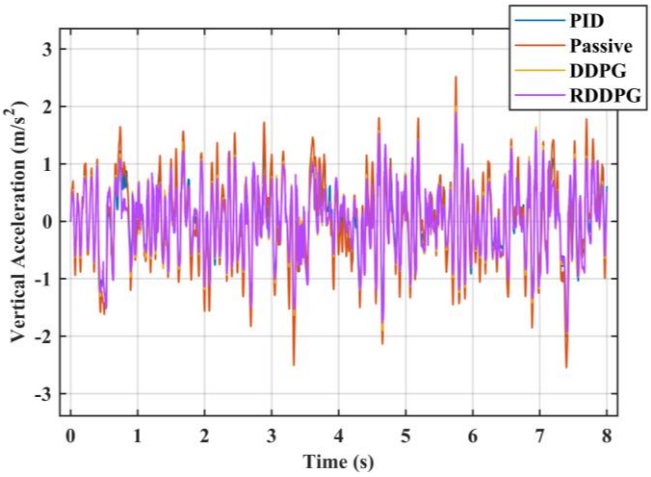

Fig. 12. Level B random road input simulation results.

TABLE VIII. SIMULATION RESULTS OF LEVEL B RANDOM ROAD INPUT CONDITION

Controller	Level B Random Road Input Condition	
	RMS	***RMS Reduction***
Passive	0.76 m/s²	-
DDPG	0.63 m/s²	17.11%
RDDPG	0.62 m/s²	18.42%
PID	0.66 m/s²	13.16%

V. DISCUSSION

Different from the residual control structure proposed by Hynes et al. [19], this study adopts an RL-centric architecture, where the DDPG controller serves as the primary decision-making module, and a PID controller is employed as a residual correction mechanism to enhance robustness. In contrast to their design, which relies on a PID-based policy and uses DDPG as a supplementary component requiring careful pre-tuning, this study approach eliminates the dependence on a well-tuned initial controller. Moreover, an Error-Aware Reward Rescaling mechanism is introduced to improve both training efficiency and policy convergence. The prior work reports results under a limited range of conditions. In contrast, this study tests the proposed method across five representative road scenarios, including both training and unseen cases, to provide a more complete assessment of its generalization and robustness.

Compared to other RL suspension control studies [20], this study demonstrates clear advantages in control architecture, adaptability, and generalization capability. First, it uses a residual control setup that combines a DDPG controller with a PID correction model. The PID component adaptively adjusts the DDPG output to enhance the stability and robustness of the control strategy under complex road conditions. An Error-Aware Reward Rescaling mechanism is introduced to adjust

reward weights based on acceleration errors during training, which improves adaptability and convergence efficiency. In terms of validation, this study evaluates the controller under five representative road profiles that include sinewaves, step disturbances, and ISO 8608 Level B random inputs. This study provides a more comprehensive assessment of generalization compared to previous work.

Simulation results consistently show that RDDPG outperforms standalone DDPG and traditional PID controllers. Under the 0.05 m sinewave excitation used as the training condition, RDDPG achieves a 50.35% reduction in RMS vertical acceleration compared to the passive baseline, exceeding the performance of DDPG at 35.02% and PID at 26.30%. Under the 0.05 m step input as an unseen condition, RDDPG still delivers the best result, reducing RMS by 20.20%, while DDPG achieves 17.17% and PID achieves 15.15%. Even in stochastic and unstructured conditions such as ISO 8608 Level B random road condition, RDDPG maintains reliable performance with the lowest RMS value of 0.62 m/s², indicating strong resilience.

Overall, by combining the long-term policy optimization ability of RL with the short-term corrective precision of PID control, the RDDPG strategy achieves a balanced performance in both stability and generalization. Its consistent effectiveness across a wide range of scenarios underscores its reliability and practical value for AS systems.

VI. Conclusion

This study adopts a hybrid control strategy for AS systems based on a residual DDPG framework. The method uses a DDPG controller together with a PID controller in a residual setup. The PID part adjusts the DDPG output to make the system more stable and robust. To further improve training efficiency and convergence, an Error-Aware Reward Rescaling strategy is introduced during the learning process. The mechanism adjusts the reward signal based on vertical acceleration deviation to guide the agent toward better control behavior. The proposed method is tested under five typical road excitations, including both training scenarios and several unseen cases. This work makes the method better by adding a way to change the reward during training. This assists to keep training stable and improves control. The results show that the better RDDPG works well in both tested and new situations, and gives a useful way to control active suspension.

VII. Future Work

Future work will try using the RDDPG method on full-vehicle suspension models to test how well it works with more axles and connected movements. Additionally, incorporating real-time implementation and validation will be considered to assess the controller's practical feasibility and robustness in real-world scenarios. Further research on alternative residual structures and adaptive weighting mechanisms may also enhance the system more flexible and work better.

Acknowledgment

This research was funded by the Ministry of Higher Education Malaysia and Universiti Kebangsaan Malaysia through FRGS research grants (Vote no: FRGS/1/2022/TK02/UKM/02/2 and a university research grant (Vote no. GGPM-2017-087), respectively.

References

[1] M. Raibail et al., "Decentralized Multi-Robot Collision Avoidance: A Systematic Review from 2015 to 2021," Symmetry, vol. 14, no. 3, p. 610, Mar. 2022, doi: 10.3390/sym14030610.

[2] N. Zulkarnain, H. Zamzuri, and S. A. Mazlan, "LQG Control Design for Vehicle Active Anti-Roll Bar System," Appl. Mech. Mater., vol. 663, pp. 146–151, Oct. 2014, doi: 10.4028/www.scientific.net/AMM.663.146.

[3] N. Binti Zulkarnain et al., "Newly Developed Nonlinear Vehicle Model for an Active Anti-roll Bar System," Bull. Electr. Eng. Inform., vol. 7, no. 4, pp. 529–537, Dec. 2018, doi: 10.11591/eei.v7i4.1185.

[4] N. Zulkarnain, A. Hussain, H. Zamzuri, and S. S. Mokri, "Controller Design for an Active Anti-roll Bar System," in 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China: IEEE, Jun. 2019, pp. 2213–2217. doi: 10.1109/ICIEA.2019.8834106.

[5] Z. Zhang, A. H. A. Rahman, N. Zulkarnain, and T. Zhang, "Enhancing Vehicle Lateral Stability: A DDPG-Based Active Anti-Roll Bar Control Strategy," IEEE Access, vol. 12, pp. 153030–153043, 2024, doi: 10.1109/ACCESS.2024.3480116.

[6] D. N. Nguyen and T. A. Nguyen, "A Hybrid Control Algorithm Fuzzy-PI with the Second Derivative of the Error Signal for an Active Suspension System," Math. Probl. Eng., vol. 2022, pp. 1–10, Aug. 2022, doi: 10.1155/2022/3525609.

[7] H. Zhou, V. Nguyen, R. Jiao, and Y. Huan, "Application of machine learning method to control the vibration of the car's suspension system," Vibroengineering PROCEDIA, vol. 38, pp. 44–49, Jun. 2021, doi: 10.21595/vp.2021.22025.

[8] S. Bongain and M. Jamett, "Electrohydraulic Active Suspension Fuzzy-Neural Based Control System," IEEE Lat. Am. Trans., vol. 16, no. 9, pp. 2454–2459, Sep. 2018, doi: 10.1109/TLA.2018.8789568.

[9] A. Konoiko, A. Kadhem, I. Saiful, N. Ghorbanian, Y. Zweiri, and M. N. Sahinkaya, "Deep learning framework for controlling an active suspension system," J. Vib. Control, vol. 25, no. 17, pp. 2316–2329, Sep. 2019, doi: 10.1177/1077546319853070.

[10] S. B. Samsuri, H. Zamzuri, M. A. A. Rahman, S. A. Mazlan, and A. H. A. Rahman, "Computational Cost Analysis of Extended Kalman Filter in Simultaneous Localization & Mapping (EKF-Slam) Problem for Autonomous Vehicle," vol. 10, no. 17, 2015.

[11] G. Liang, T. Zhao, and Y. Wei, "DDPG based self-learning active and model-constrained semi-active suspension control," in 2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI), Tianjin, China: IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/CVCI54083.2021.9661158.

[12] H. Yong, J. Seo, J. Kim, M. Kim, and J. Choi, "Suspension Control Strategies Using Switched Soft Actor-Critic Models for Real Roads," IEEE Trans. Ind. Electron., vol. 70, no. 1, pp. 824–832, Jan. 2023, doi: 10.1109/TIE.2022.3153805.

[13] A. Fares and A. Bani Younes, "Online Reinforcement Learning-Based Control of an Active Suspension System Using the Actor Critic Approach," Appl. Sci., vol. 10, no. 22, p. 8060, Nov. 2020, doi: 10.3390/app10228060.

[14] H.-C. Chen, Y.-C. Lin, and Y.-H. Chang, "An Actor-Critic Reinforcement Learning Control Approach for Discrete-Time Linear System with Uncertainty," in 2018 International Automatic Control Conference (CACS), Taoyuan: IEEE, Nov. 2018, pp. 1–5. doi: 10.1109/CACS.2018.8606740.

[15] Z. Li, T. Chu, and U. Kalabic, "Dynamics-Enabled Safe Deep Reinforcement Learning: Case Study on Active Suspension Control," in 2019 IEEE Conference on Control Technology and Applications (CCTA), Hong Kong, China: IEEE, Aug. 2019, pp. 585–591. doi: 10.1109/CCTA.2019.8920696.

[16] M.-B. Radac and T. Lala, "Learning nonlinear robust control as a data-driven zero-sum two-player game for an active suspension system,"

IFAC-Pap., vol. 53, no. 2, pp. 8057–8062, 2020, doi: 10.1016/j.ifacol.2020.12.2243.

[17] W. Wang, K. Tian, and J. Zhang, "Dynamic Modelling and Adaptive Control of Automobile Active Suspension System," J. Eur. Systèmes Autom., vol. 53, no. 2, pp. 297–303, May 2020, doi: 10.18280/jesa.530218.

[18] X. Wang, W. Zhuang, and G. Yin, "Learning-Based Vibration Control of Vehicle Active Suspension," in 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom: IEEE, Jul. 2020, pp. 94–99. doi: 10.1109/INDIN45582.2020.9442091.

[19] A. Hynes, E. Sapozhnikova, and I. Dusparic, "Optimising PID Control with Residual Policy Reinforcement Learning," presented at the 28th Irish Conference on Artificial Intelligence and Cognitive Science, Dec. 2020.

[20] A. Ahmad, M. A. Radi, A. Ahmad, and A. Boudiaf, "Reinforcement Learning-Based Optimal Control Design for Active Suspension Systems," in 2024 Advances in Science and Engineering Technology International Conferences (ASET), Abu Dhabi, United Arab Emirates: IEEE, Jun. 2024, pp. 1–6. doi: 10.1109/ASET60340.2024.10708702.