



Reinforcement learning-based control co-design of digital twin-enabled full-vehicle active suspension systems

Ying-Kuan Tsai¹ · Yi-Ping Chen¹ · Vispi Karkaria¹ · Wei Chen¹

Received: 4 November 2025 / Revised: 7 February 2026 / Accepted: 27 February 2026 / Published online: 11 April 2026
© The Author(s) 2026

Abstract

Active suspension systems are critical for enhancing vehicle comfort, safety, and stability, yet their performance is often limited by fixed hardware designs and control strategies that cannot adapt to uncertain and dynamic operating conditions. Recent advances in Digital Twins (DTs) and Reinforcement Learning (RL) offer new opportunities for real-time, data-driven optimization across a vehicle's lifecycle. However, integrating these technologies into a unified framework for co-optimizing physical and control systems remains an open challenge. This work presents an RL-based Control Co-Design (CCD) framework for full-vehicle active suspensions using multi-generation design and DT concepts. Through integrating automatic differentiation into Deep Reinforcement Learning (DRL), we jointly optimize physical components of suspension systems and control policies under varying driver behaviors and environmental uncertainties. The DRL technique also addresses the challenge of partial observability, where only limited states can be sensed and fed back to the controller, by learning optimal control actions directly from available sensor information. The framework incorporates model updating with quantile learning to quantify data uncertainty, enabling real-time decision-making and adaptive learning from digital-physical interactions. The approach demonstrates personalized optimization of autonomous suspension systems under two distinct driving settings (mild and aggressive). The results show that the optimized systems achieve smoother trajectories and reduce control efforts by approximately 58% and 12% for mild and aggressive while improving ride comfort by approximately 17% and 28%, respectively. Contributions of this work include: (1) developing a DT-enabled CCD framework integrating DRL and uncertainty-aware model updating for full-vehicle active suspensions, (2) introducing a multi-generation design framework for self-improving systems across the whole lifecycle, and (3) demonstrating personalized optimization of active suspension systems for distinct types of drivers.

Keywords Digital twin · Control co-design · Deep reinforcement learning · Full vehicle · Active suspension system · Multi-generation design · Uncertainty quantification · Real-time updating

Abbreviations

CCD	Control co-design
DNN	Deep neural network
DRL	Deep reinforcement learning
DT	Digital twin
PPO	Proximal policy optimization
RL	Reinforcement learning
UQ	Uncertainty quantification

Communicated by Zhen Hu.

✉ Wei Chen
weichen@northwestern.edu

¹ Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA

1 Introduction

1.1 Motivation

Suspension systems are central to vehicle safety, ride comfort, and handling stability (Sun et al. 2020; Goodarzi and Khajepour 2017). They directly shape how road excitations are transmitted to the cabin, influencing passenger comfort and fatigue, while also governing tire-road contact that affects braking distance, cornering capability, and overall vehicle controllability. However, traditional passive and semi-active suspensions are typically tuned for nominal operating conditions and therefore struggle under highly uncertain and variable scenarios, including diverse road roughness, tire conditions, payload changes, and heteroge-

neous driver behaviors (Soliman and Kaldas 2021; Lee et al. 2022a; Qiu et al. 2025). As operating environments deviate from the design point, these systems exhibit performance degradation, either by transmitting excessive vibrations to the occupants or by allowing excessive body motion that compromises safety and control (Tsai and Malak 2024; Weaver-Rosen et al. 2020).

To overcome these structural limitations, active suspension systems, which use actuators to actively inject energy and exert an adaptive counter-force to suppress vibrations, have been developed to provide high flexibility and improved control over vehicle dynamics. This technological shift has coincided with the emergence of learning-based control methods, particularly Deep Reinforcement Learning (DRL), which are ideally suited for addressing the complexity and nonlinearity of suspension control problems (Fares and Bani Younes 2020; Dridi et al. 2023, 2025; Ming et al. 2020; Lee et al. 2022b; Wang et al. 2024). While such approaches improve closed-loop adaptability, they remain fundamentally constrained by the fixed physical design of the suspension hardware (springs, dampers, and actuator characteristics). This motivates Control Co-Design (CCD), a design paradigm that simultaneously optimizes the physical system and the control strategy to achieve system-level optimality (Garcia-Sanz 2019; Allison and Herber 2014; Fathy et al. 2001; Cui et al. 2021; Sato et al. 2025). Existing CCD studies have demonstrated significant performance improvements compared with traditional sequential (design-then-control) approaches (Bayat and Allison 2025; Allison et al. 2014; Tsai and Malak 2026). However, they often treat environmental conditions and user behaviors as exogenous or static, and thus fail to embed these factors within the co-design loop. The result is limited adaptability when the operating context shifts.

Digital Twins (DTs) offer a path beyond these limitations by enabling real-time data collection, model updating, and predictive decision-making through a continuously synchronized virtual replica of the physical system (Semeraro et al. 2021; NASEM 2023; Thelen et al. 2022, 2023; Karkaria et al. 2024, 2025b, c; Chen et al. 2025a). Crucially, DTs also facilitate *personalized* optimization by learning the unique characteristics of individual assets and users (Hu et al. 2023; Thelen et al. 2022). In the context of vehicle suspensions, this capability is particularly critical, as uncertainties stem from both environmental variability (e.g., road profiles, friction conditions, and weather) and user-specific driving behaviors (e.g., acceleration, speed, and steering inputs). Moreover, these sources of uncertainty are inherently dynamic and evolve unpredictably over time.

To effectively handle such evolving and unstructured information, the integration of Artificial Intelligence (AI)-driven learning into DTs is highly beneficial. AI enables DTs to process heterogeneous sensor data, infer latent system

states, and continuously update predictive models, transforming the DT from a static simulation tool into an adaptive, decision-making agent. For autonomous systems such as active suspensions, this integration allows the DT to identify patterns in vehicle-road interactions and to real-time adjust the operation commands and actions. Consequently, the fusion of AI and DT technologies bridges data and design, facilitating co-optimization that enhances the adaptability, robustness, and personalization of next-generation vehicular systems (Asmat et al. 2025).

Despite this potential, most DT implementations in the suspension domain focus on monitoring and control alone (Rosa and Branco 2024; Qiu et al. 2025; Li et al. 2025), while physical design integration within the DT feedback loop remains rare (van Beek et al. 2023; Tsai et al. 2025a, b). Without coupling the hardware design with the continuously updated virtual model and data-driven control, adaptability is fundamentally limited: controllers must compensate for hardware that may be poorly matched to changing contexts, and design decisions cannot leverage the rich information provided by the in-service data. This gap motivates a DT-enabled CCD framework that unifies design, control, and learning, thereby enabling suspensions that are not only intelligent and adaptive, but also *co-evolve* physically and algorithmically over the vehicle's lifecycle.

1.2 Multi-generation digital twin concept

The concept of multi-generation design has traditionally been applied in product lifecycle engineering, remanufacturing, and sustainable design, where feedback from previous product generations informs the development of improved successors (Go et al. 2015; Karkaria et al. 2025a; van Beek et al. 2023; Asif et al. 2021; Nag et al. 2022). In conventional vehicle suspension development, this process often relies on costly prototyping campaigns and heuristic parameter tuning across limited test scenarios, making systematic exploration of design-control trade-offs challenging.

In this work, the concept is extended to the co-design of vehicle suspension systems and controllers, forming an evolving design-control ecosystem in which knowledge from each deployment cycle is systematically assimilated into the DT and used to guide subsequent redesigns. By leveraging self-learning models, learning-based co-design, and virtual evaluation environments, the DT enables efficient assessment of physical and control design alternatives under diverse and uncertain operating conditions, significantly reducing reliance on repeated physical experimentation. Through this mechanism, the DT serves as a persistent learning agent that accumulates operational knowledge over time, continually improving its predictive accuracy and enabling progressively better physical and control designs across generations.

Our prior work (Tsai et al. 2025a) demonstrated that the integration of DTs with CCD and DRL can significantly enhance the adaptability of active suspension systems. In particular, a quarter-car suspension case study showed that the proposed multi-generation DT framework enables the system to become more robust to varying initial conditions and uncertain road roughness. By continuously updating the digital model and co-optimizing the physical parameters and control policy, each successive generation achieved superior performance under broader uncertainty conditions. However, the previous study only considered a simplified suspension configuration that did not capture the complex interactions and couplings present in a full-vehicle system, and motivates us to work on more realistic, complex, and scaled testbeds.

Extending the framework to a *full-vehicle* active suspension system introduces several new challenges that necessitate further methodological developments. First, the system is partially observable, meaning that not all states are directly measurable from onboard sensors. This requires the policy to infer unobserved dynamics through learned correlations. Second, the problem involves a high-dimensional state and action space, as the full vehicle includes coupled lift, pitch, and roll motions with four actuators that must be coordinated in real time. Third, the optimization must ensure system-level dynamic controllability across varying driving and environmental conditions, balancing comfort, stability, and energy efficiency under multiple operating modes.

1.3 Research objectives and contributions

By addressing the challenges of full-vehicle modeling, partial observability, coupled dynamics with proactive control, the DT framework extends beyond traditional offline control design toward a data-driven, learning-based, and self-improving paradigm. Through iterative model updating, uncertainty quantification (UQ), and RL-based co-optimization, the framework continuously redesigns both the physical system and its controller for future generations, enabling adaptive, personalized, and sustainable vehicle performance throughout the system lifecycle.

The objective of this work is to present a multi-generation DT framework to co-optimize the physical components and active controllers for full-vehicle active suspension systems by integrating automatic differentiation with CCD and DRL techniques. This framework also enables data-informed adaptation of suspension design and control strategies by leveraging real-time monitoring, model updating, and decision-making within digital twins, allowing performance to be evaluated and improved across a wide range of operating scenarios.

This research makes the following key contributions:

- We propose a DT-enabled CCD methodology that jointly optimizes full-car suspension components and learning-based control policies within a unified DRL framework.
- We use a multi-generation design framework integrated with CCD formulations and the DT technology to enable continuous adaptation and performance improvement across successive generations.
- We demonstrate how the proposed framework can be used to analyze suspension design–control trade-offs under distinct driving styles and operating conditions (e.g., mild versus aggressive driving), providing design insight that supports robust tuning and future-generation design decisions.

This work establishes a pathway toward more intelligent and adaptive DTs that support decisions on suspension design and control development across the vehicle lifecycle and improve ride comfort, robustness, and energy efficiency under diverse operating conditions and evolving usage profiles. Consistent with recent digital twin literature, we view a digital twin as a progressive cyber–physical system that evolves from simulation-based initialization to experimental integration and real-time updating; the present study focuses on this simulation-based initialization stage.

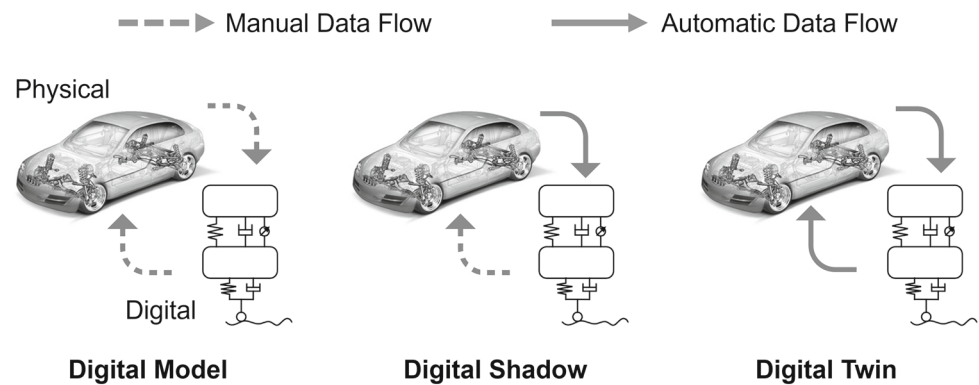
The remainder of this article is organized as follows. Section 2 reviews the technical background on DTs, CCD formulations, and RL techniques. Section 3 formulates the full-vehicle active suspension dynamics, describes the numerical solution procedure, models external disturbances, and states the problem definition. Section 4 details the multi-generation CCD framework with the integration of DT and DRL, and its stepwise implementation. Section 5 presents results and extended discussion. Finally, Sect. 6 concludes the paper and outlines directions for future research.

2 Background

2.1 Digital twin

A Digital Twin (DT) is generally defined as a virtual replica of a physical system that is continuously updated with data collected from the physical asset throughout its life cycle (Grieves 2014; Thelen et al. 2022, 2023; Karkaria et al. 2025b). Unlike traditional modeling approaches that typically focus on a single design stage or a single generation of a product, a DT spans the entire product life cycle, from conceptual design (Wang et al. 2020) and manufacturing (Onaji et al. 2022) to operation (Yu et al. 2021), maintenance (Karkaria et al. 2025a; Hu et al. 2024), and next-generation redesign (van Beek et al. 2023). This continuous and bi-directional connection allows the DT not only to mirror the physical system but also to inform decision-making

Fig. 1 Illustration of the differences among a digital model, a digital shadow, and a digital twin



(Chen et al. 2025a) and enable adaptive improvements over time.

It is important to clarify the distinction between a digital model, a digital shadow, and a DT, as these terms are often used interchangeably in the literature (Grieves and Vickers 2017). Figure 1 illustrates the three levels of digital representations using the example of an active suspension system. A digital model refers to a static virtual model of the physical system that is created during the design phase. This model is not automatically updated with real-world data. For example, a vehicle manufacturer may simulate suspension performance under pre-defined road profiles, but the model remains unchanged once it is built. A digital shadow, in contrast, is a virtual representation of the physical counterpart that receives a one-way flow of data from the physical system or real environment. Sensor measurements are streamed into the model, allowing it to reflect the current state of the physical system. However, the shadow does not send information back to the physical system. That is, no decision-support or feedback actions are transmitted from the digital to the physical domain. Finally, a DT represents a fully connected virtual replica of the physical system with bi-directional data flow. The twin not only updates its states with real-time sensor data but also uses predictive models and makes informed decisions for the real system. This closed-loop interaction enables the system to adapt to varying conditions and changing environments, thereby enhancing performance, robustness, and adaptability. In this work, the digital twin is considered at a pre-deployment stage, where the physical system is represented by a surrogate model that emulates unmodeled nonlinearities, disturbances, and model–reality mismatch prior to real-world implementation.

It is worth noting that in many practical cases, the bi-directional communication within a DT may not be fully automatic but instead involve human decision-makers in the loop (Thelen et al. 2022; Sisson et al. 2022; Marykovskiy et al. 2024). For example, predictive maintenance applications often rely on the DT to continuously monitor sensor data, identify early signs of component degradation, and recommend maintenance actions (Karkaria et al. 2025a).

Although the final decision and execution may require human approval, these actions must be handled in a timely manner to prevent unexpected breakdowns and ensure system reliability. Similarly, in remanufacturing scenarios, a DT can assess the condition of returned components, estimate their remaining useful life, and suggest refurbishment strategies, with human operators executing the recommended actions (Karkaria et al. 2025c). In such cases, the DT functions as an intelligent decision-support system, ensuring that feedback between the physical and digital entities is closed effectively even when human involvement is necessary.

2.2 Control co-design

Control Co-Design (CCD) is an integrated framework that concurrently addresses the design of physical systems (plants) and their control strategies, accounting for the interdependence between system dynamics and control behavior. Unlike conventional sequential approaches, CCD enables designers to achieve superior system-level performance by jointly considering both domains (Garcia-Sanz 2019; Allison and Herber 2014; Tsai and Malak 2022a; van Diepen and Shea 2022). Two prevalent CCD paradigms exist: simultaneous and nested formulations (Herber and Allison 2019). The simultaneous approach treats physical and control variables within a unified optimization problem, whereas the nested approach decomposes the problem hierarchically, with the outer loop handling physical design and the inner loop solving for control under fixed physical parameters. The nested strategy is particularly advantageous when distinct objectives govern the control and physical design, or when the control synthesis relies on domain-specific methods (Nash et al. 2021; Tsai and Malak 2023, 2025).

The choice between these formulations is guided by factors such as problem structure, computational considerations, and the tools available for controller synthesis. In this study, we adopt a simultaneous CCD framework, which enables co-optimization of the physical components and control policy through Deep Reinforcement Learning (DRL) with the integration of automatic differentiation. This learning-based

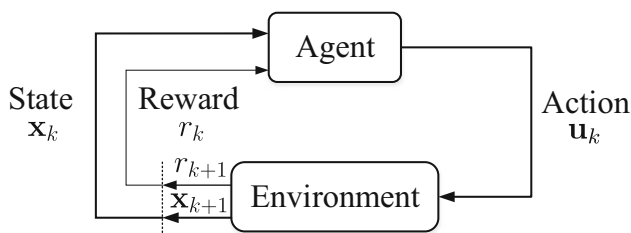


Fig. 2 Diagram of Reinforcement Learning (RL), modified from Sutton and Barto (1998)

method captures the complex interdependencies between design and control and facilitates system adaptability under dynamic and uncertain conditions by leveraging real-time data.

2.3 Reinforcement learning

Reinforcement Learning (RL) provides a mathematical framework for solving sequential decision-making problems, where an agent learns to make decisions through interactions with its environment in order to maximize long-term rewards (Buşoniu et al. 2018), as illustrated in Fig. 2. This learning process is typically modeled using a Markov Decision Process (MDP), represented by the tuple $(\mathcal{X}, \mathcal{U}, P, R, \gamma)$. Here, \mathcal{X} denotes the state space, \mathcal{U} the action (or control input) space, P the system’s transition model, R the reward function, which is often interpreted as the negative of a cost, and $\gamma \in (0, 1]$ the discount factor. In an MDP framework, state transitions are generally stochastic, meaning that applying an action \mathbf{u}_k to a system in state \mathbf{x}_k at time step k results in a probabilistic next state \mathbf{x}_{k+1} . The transition dynamics are captured by the probability distribution $P(\mathbf{x}_{k+1}|\mathbf{x}_k, \mathbf{u}_k)$, describing the likelihood of moving to state \mathbf{x}_{k+1} from \mathbf{x}_k under action \mathbf{u}_k . In control-oriented contexts, however, it is often more practical to express the system’s evolution deterministically as a function of state and control action:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \tag{1}$$

which is a stochastic process according to the transition function $P(\mathbf{x}_{k+1}|\mathbf{x}_k, \mathbf{u}_k)$.

After transitioning to the next state \mathbf{x}_{k+1} , the agent receives a scalar reward defined as $r_{k+1} = R(\mathbf{x}_k, \mathbf{u}_k, \mathbf{x}_{k+1})$, where the reward function $R : \mathcal{X} \times \mathcal{U} \times \mathcal{X} \rightarrow \mathbb{R}$ maps the current state, action, and resulting state to a real-valued reward. The agent’s behavior is governed by a policy, which is a probabilistic mapping from states to actions. Specifically, given a current state \mathbf{x}_k , the policy π assigns probabilities to possible actions such that the agent selects \mathbf{u}_k according to the distribution $\pi(\mathbf{u}_k|\mathbf{x}_k)$.

The objective in RL is to maximize the return, defined as the cumulative discounted sum of future rewards starting from time step k :

$$G_k = \sum_{i=0}^{\infty} \gamma^i r_{k+i+1}, \tag{2}$$

where $\gamma \in (0, 1]$ is the discount factor, which determines the importance of future rewards relative to immediate ones.

To evaluate how good a state is under a given policy π , we define the state-value function:

$$\begin{aligned} V^\pi(\mathbf{x}) &= \mathbb{E}_\pi [G_k | \mathbf{x}_k = \mathbf{x}] \\ &= \mathbb{E}_\pi \left[\sum_{i=0}^{\infty} \gamma^i R(\mathbf{x}_{k+i}, \mathbf{u}_{k+i}, \mathbf{x}_{k+i+1}) | \mathbf{x}_k = \mathbf{x} \right], \end{aligned} \tag{3}$$

for all $\mathbf{x} \in \mathcal{X}$. This function, also called the V-function, estimates the expected return when starting from state \mathbf{x} and following the policy π thereafter.

While the state-value function captures the quality of states under a policy, effective decision-making requires evaluating the quality of specific actions taken from a given state. This leads to the definition of the action-value function, or Q-function, as:

$$\begin{aligned} Q^\pi(\mathbf{x}, \mathbf{u}) &= \mathbb{E}_\pi [G_k | \mathbf{x}_k = \mathbf{x}, \mathbf{u}_k = \mathbf{u}] \\ &= \mathbb{E}_\pi \left[\sum_{i=0}^{\infty} \gamma^i R(\mathbf{x}_{k+i}, \mathbf{u}_{k+i}, \mathbf{x}_{k+i+1}) | \mathbf{x}_k = \mathbf{x}, \mathbf{u}_k = \mathbf{u} \right], \end{aligned} \tag{4}$$

for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{u} \in \mathcal{U}$. The Q-function estimates the expected return of executing action \mathbf{u} in state \mathbf{x} and then continuing with policy π thereafter.

Conventional RL approaches, such as tabular Q-learning and dynamic programming, face significant challenges when dealing with complex control tasks involving continuous state and action spaces. In these scenarios, it is infeasible to explicitly represent or update the value function for all possible state-action combinations, as the space becomes effectively infinite. This scalability issue renders traditional RL methods unsuitable for high-dimensional control applications, including robotic manipulation, autonomous vehicle control, and real-time decision-making in DT-enabled systems.

Deep Reinforcement Learning (DRL) overcomes the limitations of traditional RL methods in continuous and high-dimensional spaces by employing Deep Neural Networks (DNNs) to approximate value functions, policies, or both Buşoniu et al. (2018); Shakya et al. (2023). Rather than relying on explicit tables to store state-action values, DRL

leverages the generalization capability of neural networks to estimate functions over large or continuous domains. This approach allows DRL to scale effectively to complex environments. Among the popular DRL algorithms are policy-based methods such as Trust Region Policy Optimization (TRPO) (Schulman et al. 2015) and Proximal Policy Optimization (PPO) (Schulman et al. 2017), which improve learning stability and efficiency through constrained updates. These methods ensure that policy changes remain within a “trust region,” thereby avoiding abrupt shifts that could hinder convergence. Beyond these, numerous DRL algorithms have been developed to address challenges like sample efficiency, exploration-exploitation trade-offs, and training stability. For broader overviews of the DRL landscape, readers may consult comprehensive surveys in Arulkumaran et al. (2017); Ladosz et al. (2022); Wang et al. (2022). The integration of deep learning into RL has dramatically expanded its practical relevance, enabling application to real-world systems where classical approaches fall short due to computational constraints.

2.4 Co-design using reinforcement learning

In the CCD literature, physical design parameters are consistently treated as a distinct class of plant design variables, rather than as components of the system state. These variables (e.g., stiffness, damping, geometry, mass properties, and actuator or sensor placement) define the physical configuration of the system and remain fixed during operation, while the system state describes the transient dynamical behavior governed by this configuration. This distinction is emphasized in integrated control co-design formulations, where plant design variables and control variables are jointly optimized but are not included in the state vector (Allison and Herber 2014; Nash et al. 2021; Jin and Schmit 1992). Although such parameters can be mathematically embedded as states with zero dynamics, this representation is generally not adopted in co-design formulations, as it obscures their role as persistent, deployable design decisions rather than evolving system behavior.

Given the remarkable success of DRL in solving complex control problems, its integration into co-design frameworks has attracted significant attention in recent years. In DRL-based co-design, the goal is to jointly optimize both the physical parameters of a system and its control policy, enabling a unified treatment of structural design and decision-making under uncertainty (Ma et al. 2021; Yuhn et al. 2023; Chen et al. 2020; Sun et al. 2023; Schaff et al. 2019; Luck et al. 2020; Wang et al. 2023). This paradigm reflects a shift from the traditional sequential design-then-control process toward an adaptive, closed-loop optimization that allows physical morphology and control behavior to evolve together. Such co-adaptive mechanisms

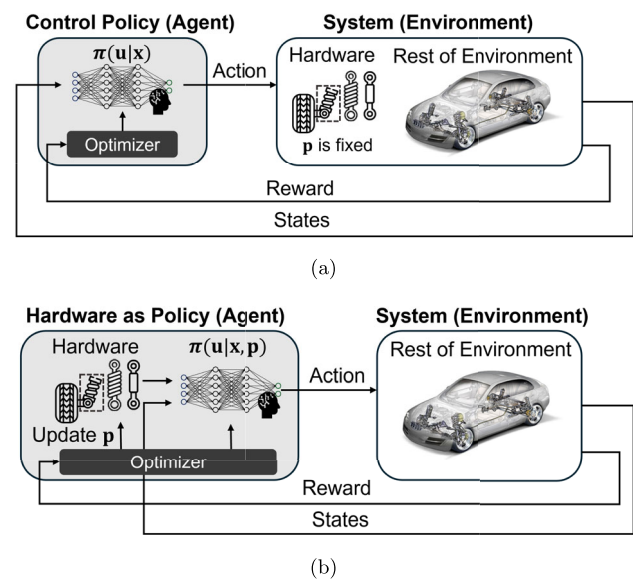


Fig. 3 Comparison of **a** conventional DRL-based control with fixed hardware parameter p , where the policy π is only dependent of p , and **b** hardware as policy, where the policy π is dependent of x and p and the hardware parameter and the policy are co-optimized (modified from Chen et al. (2020))

have shown improved robustness, adaptability, and performance in dynamic environments, particularly in robotic applications such as manipulators, legged locomotion, and modular soft robots.

A common strategy for solving learning-based co-design problems is to employ a bi-level optimization architecture, in which the inner loop trains a control policy for a fixed physical configuration, while the outer loop searches the design space through methods such as evolutionary algorithms or Bayesian optimization (Schaff et al. 2019; Wang et al. 2019; Gupta et al. 2021; Chen et al. 2023b). This separation allows each level of problem to use specific solvers: RL for control and black-box search for system design which is suitable for discrete or non-differentiable hardware spaces. However, the approach is often computationally intensive because each modification to the physical design requires retraining the controller from scratch, resulting in high sample cost and slow convergence (Chen et al. 2023a).

To improve learning efficiency, an emerging line of research treats the mechanical design as part of the policy network itself, a concept known as hardware as policy (Chen et al. 2020; Sun et al. 2023). In this formulation, physical parameters are represented as differentiable components within the computational graph, allowing gradient-based DRL algorithms (e.g., PPO or TRPO) to co-optimize both neural-network weights and design variables through back-propagation (Schaff et al. 2019; Luck et al. 2020; Ma et al. 2021). Figure 3 illustrates this concept by contrasting (a) conventional DRL-based control, where the hardware

parameters \mathbf{p} are fixed and the policy $\pi(\mathbf{u}|\mathbf{x})$ is learned independently, with (b) the hardware-as-policy paradigm, in which both \mathbf{p} and the policy $\pi(\mathbf{u}|\mathbf{x}, \mathbf{p})$ are jointly optimized through a shared computational graph. This simultaneous optimization significantly improves sample efficiency and coordination between morphology and control, producing systems that exhibit natural co-adaptation analogous to biological evolution. Despite their success, these methods typically assume stationary training environments and do not explicitly account for the stochastic, nonstationary conditions common in real engineering systems.

To address this gap, recent efforts have extended DRL-based co-design frameworks to engineering domains beyond robotics, where the physical dynamics are more complex, the environment is less predictable, and data collection is costly (He and Su 2023; Xu and Carrillo 2017). Our previous work (Tsai et al. 2025a) represents one of the earliest applications of DRL-based CCD beyond robotics by demonstrating its use in a quarter-car active suspension system. That study introduced a gradient-based CCD formulation in which the physical suspension parameters and control policy were jointly optimized via automatic differentiation, enabling adaptive learning from road disturbances. Building on that foundation, the present study generalizes the framework to a full-vehicle active suspension system to facilitate real-time DT model and DRL policy update, co-optimize the physical system and controller, and improve the performance and adaptability across multiple generations.

While the physical design parameters optimized in such co-design formulations are fixed after deployment and do not adapt online, RL plays a critical role in determining these parameters based on expected closed-loop performance under uncertainty, rather than through open-loop or simplified surrogate objectives. In DRL-based co-design, candidate plant designs are evaluated through their interaction with the control policy across distributions of operating conditions and disturbances, allowing nonlinear dynamics, stochastic excitations, control constraints, and performance trade-offs to be naturally embedded in the optimization process. From this perspective, RL serves as an effective offline co-design mechanism, complementing real-time control adaptation by enabling plant parameters to be selected in a manner that is robust to uncertainty and tightly coupled with closed-loop behavior, which is difficult to achieve using conventional optimization methods that decouple plant design from control synthesis.

3 Simulation of full-vehicle active suspension

Active suspension systems, unlike passive suspension systems, significantly enhance ride comfort and vehicle stability

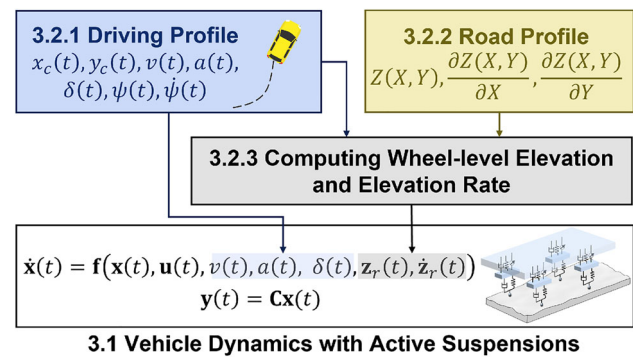


Fig. 4 Simulation workflow for the full-vehicle active suspension system, where x_c and y_c denote the longitudinal and lateral positions of the vehicle center of mass, respectively. The driving and road profiles are combined to compute wheel-level elevations and elevation rates, which serve as external disturbances to the dynamic model

by actively modulating suspension forces in real time. Designing such systems demands a close integration of physical and control domains to ensure robust performance across diverse road and driving conditions. While traditional proportional–integral–derivative (PID) controllers and linear-quadratic regulators (LQR) are widely used due to their simplicity (Anh 2020; Nguyen and Nguyen 2023; Manna et al. 2022), they lack adaptability and require meticulous manual tuning, limiting their effectiveness in dynamic environments. While our prior work demonstrated that CCD of an active suspension system across multiple generations leads to improved performance and robustness (Tsai et al. 2025a), it was limited to a quarter-car suspension model, which oversimplifies the complexity of real-world vehicle dynamics.

This section demonstrates the modeling and simulation of the full-vehicle active suspension system used in the case study. The objective is to illustrate how the vehicle dynamics are formulated and simulated under realistic operating conditions. The full-vehicle model serves as a digital representation that interacts with the DRL-based controller. We will use the simulation environment to show how the integration of real-time data collection and DT model updates enhances the system's adaptability and dynamic performance (including driving quality, vehicle stability, and passenger comfort). Additionally, we will use the simulation environment to illustrate how the framework identifies distinct optimal solutions tailored to two distinct driving behaviors (mild and aggressive drivers) for personalized optimization and improving system robustness in real-world scenarios.

Figure 4 illustrates the overall simulation flow for the full-vehicle active suspension system. The simulation begins by defining the driving profile, which specifies the vehicle's longitudinal and lateral motion, including position, speed, acceleration, steering angle, and yaw dynamics. In parallel, the road profile describes the terrain elevation and its

spatial gradients, capturing variations in road roughness and slope. These two components are combined to compute the wheel-level elevation and elevation rate, which act as external disturbances to the suspension system. The computed disturbances, together with the vehicle’s dynamic states and control inputs, form the inputs to the nonlinear state-space model of the vehicle dynamics. The model then predicts the time evolution of the system states and observable outputs, providing a realistic representation of full-vehicle behavior under diverse driving and road conditions.

3.1 Dynamic modeling

The vehicle dynamics with active suspension systems are represented as a nonlinear state-space system:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), v(t), a(t), \delta(t), \mathbf{z}_r(t), \dot{\mathbf{z}}_r(t)), \quad (5)$$

where $\mathbf{x}(t)$ is the system state vector, defined by:

$$\mathbf{x}(t) = [z_s(t), \alpha(t), \beta(t), z_{u1}(t), z_{u2}(t), z_{u3}(t), z_{u4}(t), \dot{z}_s(t), \dot{\alpha}(t), \dot{\beta}(t), \dot{z}_{u1}(t), \dot{z}_{u2}(t), \dot{z}_{u3}(t), \dot{z}_{u4}(t)]^T, \quad (6)$$

$\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t), u_4(t)]^T$ is the vector of actuator control forces applied to the four suspension systems, $v(t)$ and $a(t)$ denote the vehicle’s longitudinal velocity and acceleration, $\delta(t)$ is the steering angle, and $\mathbf{z}_r(t) = [z_{r1}(t), z_{r2}(t), z_{r3}(t), z_{r4}(t)]^T$, $\dot{\mathbf{z}}_r(t) = [\dot{z}_{r1}(t), \dot{z}_{r2}(t), \dot{z}_{r3}(t), \dot{z}_{r4}(t)]^T$ are the vectors of road heights and road velocity inputs at four wheels.

The dynamic equations of sprung mass vertical motion, pitch motion, roll motion, and unsprung mass (wheel) vertical motion, respectively, are:

$$m_s \ddot{z}_s(t) = \sum_{i=1}^4 (F_{si}(t) + u_i(t)), \quad (7)$$

$$I_\alpha(t) \ddot{\alpha} = \sum_{i=1}^4 (F_{si}(t) + u_i(t)) d_{xi} + M_\alpha(t), \quad (8)$$

$$I_\beta(t) \ddot{\beta} = \sum_{i=1}^4 (F_{si}(t) + u_i(t)) d_{yi} + M_\beta(t), \quad (9)$$

$$m_{ui} \ddot{z}_{ui}(t) = F_{ti}(t) - F_{si}(t) - u_i(t), \quad \forall i \in \{1, 2, 3, 4\}, \quad (10)$$

where $F_{si}(t)$ denotes the suspension force at Wheel i , defined by:

$$F_{si}(t) = k_s(z_{ui}(t) - z_s(t) - \Delta_i(t)) + c_s(\dot{z}_{ui}(t) - \dot{z}_s(t) - \dot{\Delta}_i(t)), \quad (11)$$

and $F_{ti}(t)$ denotes the tire (contact) force at Wheel i , defined by:

$$F_{ti}(t) = k_t(z_{ri}(t) - z_{ui}(t)) + c_t(\dot{z}_{ri}(t) - \dot{z}_{ui}(t)), \quad (12)$$

where Δ_i accounts for the geometric displacement due to pitch and roll at each wheel location:

$$\Delta_1(t) = -l_f \alpha(t) + \frac{l}{2} \beta(t), \quad (13)$$

$$\Delta_2(t) = -l_f \alpha(t) - \frac{l}{2} \beta(t), \quad (14)$$

$$\Delta_3(t) = l_f \alpha(t) + \frac{l}{2} \beta(t), \quad (15)$$

$$\Delta_4(t) = l_f \alpha(t) - \frac{l}{2} \beta(t), \quad (16)$$

$d_{xi} = \{-l_f, -l_f, l_r, l_r\}$ and $d_{yi} = \{l/2, -l/2, l/2, -l/2\}$ denote the longitudinal and lateral distances, respectively, from the vehicle’s center of gravity to each wheel, k_t and c_t are tire stiffness and damping constant, respectively, and k_s and c_s are the coefficients of the spring and the damper (design variables for the physical system). The additional torques $M_\alpha(t)$ and $M_\beta(t)$ are generated by inertial coupling due to longitudinal and lateral accelerations (e.g., braking and cornering), modeled as $M_\alpha(t) = m_s h_{CG} a(t)$ and $M_\beta(t) = m_s h_{CG} \frac{v(t)^2 \tan \delta(t)}{l_f + l_r}$, with h_{CG} representing the height of the vehicle’s center of gravity (CG) above ground. The physical and geometric parameters used in the full-vehicle dynamic model are summarized in Table 1.

This model captures seven degrees of freedom (DOFs): three DOFs for the sprung mass (lift z_s , pitch α , and roll β), and four vertical DOFs for the unsprung masses (wheels) $z_{u1} - z_{u4}$, shown in Fig. 5. We assume small angles for pitch and roll to simplify the rotational dynamics. Tire-road interactions are modeled with vertical spring-damper pairs, and control forces $\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t), u_4(t)]^T$ are treated as external actuator inputs that directly adjust the actuation forces for the four suspension systems.

Suspension systems often suffer from limited observability due to the difficulty of directly measuring all internal states in real-time, especially the motion and dynamics of the sprung body. In full-vehicle suspension models, the system dynamics typically involve a high-dimensional state vector, while only a subset of states can be accessed through physically realizable sensors. Accordingly, the observed state vector $\mathbf{y}(t)$ is selected to balance practical sensor availability, robustness to noise, and relevance to ride comfort and handling objectives, rather than to achieve full-state observability. In this system, the observed state vector $\mathbf{y}(t)$ is defined

Table 1 Vehicle parameters

Symbol	Physical meaning	Value	Unit
m_s	Sprung mass	1500	kg
I_α	Pitch rotation inertia	2500	kg m ²
I_β	Roll rotation inertia	500	kg m ²
m_{us}	Unsprung mass	50	kg
k_t	Stiffness of tire	200000	N/m
c_t	Damping of tire	150	N s/m
l_f	Distance between CG and front axles	1.35	m
l_r	Distance between CG and rear axles	1.35	m
l	Track front, track rear	0.75	m
h_{CG}	Height of the CG above the ground	0.55	m

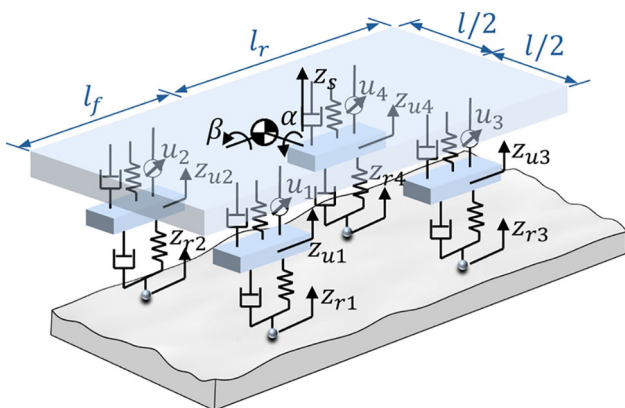


Fig. 5 Model of full vehicle with active suspension systems

by:

$$\begin{aligned}
 \mathbf{y}(t) = & [\dot{z}_s(t), \dot{\alpha}(t), \dot{\beta}(t), \\
 & z_{u1}(t), z_{u2}(t), z_{u3}(t), z_{u4}(t), \\
 & z_s(t) - z_{u1}(t), z_s(t) - z_{u2}(t), \\
 & z_s(t) - z_{u3}(t), z_s(t) - z_{u4}(t)]^\top \\
 = & \mathbf{C}\mathbf{x}(t),
 \end{aligned}
 \tag{17}$$

where $\mathbf{C} \in \mathbb{R}^{11 \times 14}$ is the observation matrix, allowing us to extract the following 11 measurable outputs from the full system. This measurement set reflects sensing modalities that are commonly available or feasible in real vehicle implementations. The vertical velocity of the vehicle body $\dot{z}_s(t)$, and pitch rate $\dot{\alpha}(t)$ and roll rate $\dot{\beta}(t)$ can be obtained from a gyroscope module. The vertical positions of the unsprung masses $z_{ui}(t)$, $i = 1, \dots, 4$, are measured via linear potentiometers attached to each wheel. Additionally, the relative displacement between the body and wheels $z_s(t) - z_{ui}(t)$ is measured using suspension stroke sensors. We note that the measurement vector is not intended to represent a minimal sensing configuration. While some relative suspension displacements can be reconstructed algebraically from wheel

position measurements and a single relative displacement, they are included here as part of the observation vector to explicitly expose suspension deflection information to the learning-based controller. This formulation should be interpreted as a modeling abstraction of available signals rather than a prescription of required physical sensors.

The controller is designed to operate under partial state measurement, consistent with practical sensing limitations and learning-based control formulations. Rather than relying on full-state observability or explicit state reconstruction, the DRL policy directly maps the available measurements to control actions, with performance validated empirically through closed-loop simulations under stochastic disturbances.

To implement the dynamic model into the DRL framework where the agent interacts with the environment at discrete-time intervals, the continuous-time vehicle dynamics are discretized using the fourth-order Runge–Kutta (RK4) method. Given the current system state $\mathbf{x}(t)$ and control input $\mathbf{u}(t)$, the RK4 method computes the next state \mathbf{x}_{k+1} over a time step $\Delta t = 0.01$ second by evaluating intermediate derivatives of the dynamics function \mathbf{f} , where k represents the index of the time step. The selected time step is sufficiently small relative to the dominant dynamics of the vehicle suspension system, ensuring adequate numerical accuracy and stability. A sampling time of 10 ms is also consistent with typical sensing and control update rates used in automotive suspension systems, and preliminary tests with smaller time steps showed similar closed-loop performance trends.

With the prescribed Δt , the dynamics can be approximated and Eq. (5) can be written in a discrete-time version:

$$\mathbf{x}_{k+1} = \mathbf{f}_d(\mathbf{x}_k, \mathbf{u}_k, v_k, a_k, \delta_k, \mathbf{z}_{rk}, \dot{\mathbf{z}}_{rk}), \tag{18}$$

which serves as the transition model in the agent–environment interaction loop. It is noted that the full system state \mathbf{x}_k is not directly observable in practice. The agent instead receives a partially observed state (or observation) vector \mathbf{y}_k . The DRL policy $\pi(\mathbf{u}_k | \mathbf{y}_k)$ is therefore trained on the observation space rather than the full state, reflecting the realistic

feedback structure in which the controller operates using only sensor-accessible information.

The vehicle dynamic model used in this study is based on several simplifying assumptions. First, it is assumed that all four wheels maintain continuous contact with the road surface, i.e., no wheel lift-off or loss of ground contact occurs. The road excitations are applied only in the vertical direction, and lateral or longitudinal tire slip dynamics are not considered. In addition, internal disturbances such as drivetrain vibrations, engine torque fluctuations, or transmission dynamics are neglected (Dridi et al. 2025). The suspension elements are assumed to be ideal, with no friction, backlash, or hysteresis effects, and the actuators apply control forces directly without delay or saturation. Sensor measurements are assumed to be noise-free and perfectly accurate. While real-world vehicle systems inevitably involve sensor noise, bias, and delays, this assumption is adopted to decouple sensing uncertainty from the core investigation of CCD and closed-loop performance under stochastic environmental disturbances, such as randomized road excitations, speed perturbations, and steering noise.

This full-vehicle suspension model presents several challenges for control:

- High dimensionality and coupled dynamics,
- Partial observability (not all states may be measurable in practice), i.e., only \mathbf{y}_k is observable instead of \mathbf{x}_k at time step k ,
- Highly uncertain environments, such as uneven terrain and aggressive driving maneuvers, and
- The need for rapid and adaptive decision-making to maintain vehicle stability and passenger comfort.

DRL updates the control policy in real time by learning from physical data to adapt to dynamic environments. Furthermore, DRL is well-suited for systems where the cost of modeling uncertainty and real-time feedback is high, and where traditional control strategies may struggle to maintain performance under unpredictable driving conditions.

3.2 External disturbance

3.2.1 Driving profiles

To demonstrate the capability of the proposed DT-enabled CCD framework in tailoring optimal suspension system designs for different user behaviors, we developed two contrasting driving profiles representative of mild and aggressive drivers. These profiles serve as external excitation scenarios, enabling evaluation of vehicle performance and adaptation under varied longitudinal and lateral dynamics.

The mild driver is characterized by gradual acceleration, moderate steering actions, and a lower cruising speed. In

contrast, the aggressive driver exhibits rapid acceleration, more abrupt steering transitions, and a higher cruising speed. These behavioral patterns are embedded in the acceleration and steering angle trajectories, which serve as inputs for vehicle motion integration.

We constructed both profiles over a duration of 1,200 s with a step of 0.01 s (a total of 120,000 steps). The longitudinal acceleration profile $a(t)$ was divided into three phases: acceleration, cruising, and braking. The mild driver accelerates at 2.0 m/s^2 to a cruising speed of 12 m/s, while the aggressive driver reaches 20 m/s under a higher acceleration of 6.0 m/s^2 (shown in Fig. 6).

Lateral control is represented by scheduled steering angle inputs, modeled using piecewise-defined angular perturbations repeated over the full simulation. The aggressive driver performs frequent and sharp turns (up to approximately 0.17 rad), while the mild driver exhibits infrequent and smoother maneuvers (within around 0.05 rad), shown as Fig. 7.

All profiles were smoothed using a Savitzky-Golay filter (Schmid et al. 2022) to emulate realistic driver commands while preserving sharp dynamic features. The resulting position (x, y) , velocity v , yaw ψ , and yaw rate $\dot{\psi}$ define the vehicle's trajectory across the road surface, from which the spatially varying road elevations are queried at each wheel as a function of time. This ensures that the excitation inputs applied to the four suspension systems correspond accurately to the instantaneous contact locations of the tires as the vehicle moves along the road, which will be detailed in Sect. 3.2.3.

Figures 6 and 7 illustrate selected segments of the longitudinal acceleration and steering angle profiles, respectively, highlighting the key differences between the driving behaviors. The aggressive profile features pronounced peaks in acceleration and steering, leading to sharper velocity and orientation changes. The mild profile, in contrast, emphasizes smoothness and gradual transitions. These user-specific trajectories are used as input scenarios to evaluate the adaptability of the suspension design under different driving styles.

3.2.2 Road profile

To evaluate the performance of the full-vehicle active suspension system under realistic conditions, we generated a high-fidelity two-dimensional (2D) road surface profile $Z(X, Y)$ that captures both micro-scale stochastic roughness and macro-scale topographic features. This approach goes beyond traditional one-dimensional bump or cleat models used in (Dridi et al. 2025) by introducing lateral variability and a spatially continuous field suitable for evaluating vehicle responses across all four wheels simultaneously.

The 2D road roughness profile was synthesized using inverse Fourier transformation of a Hermitian-symmetric complex spectrum constructed according to the ISO 8608

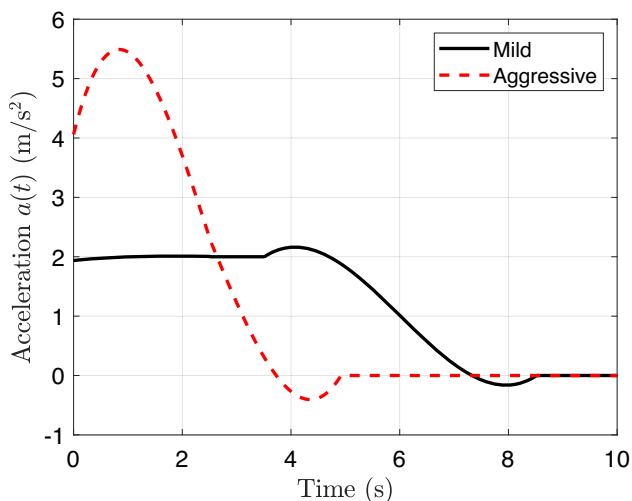


Fig. 6 Comparison of longitudinal acceleration profiles for mild and aggressive drivers. The aggressive driver accelerates more quickly and reaches a higher cruising speed

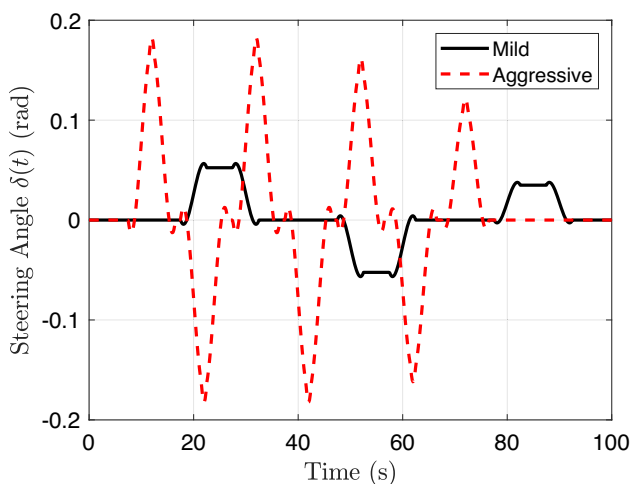


Fig. 7 Steering angle trajectories for mild and aggressive drivers. The aggressive driver demonstrates higher frequency and amplitude of steering commands

standard. The power spectral density (PSD) was defined as:

$$\Phi(n_X, n_Y) = S_0 \left(\frac{n_0}{\sqrt{n_X^2 + n_Y^2 + \epsilon}} \right)^\omega \tag{19}$$

where S_0 is the reference PSD value (e.g., $1 \times 10^{-4} \text{ m}^3$), n_X and n_Y represent the numbers of grids in x and y directions, respectively, $n_0 = 0.1$ cycles/m is the reference spatial frequency, and $\omega = 2.5$ is the waviness exponent. A small constant $\epsilon = 0.005$ m was introduced to avoid division by zero. To ensure a real-valued elevation field, Hermitian symmetry was enforced on the frequency domain matrix prior to applying the inverse FFT.

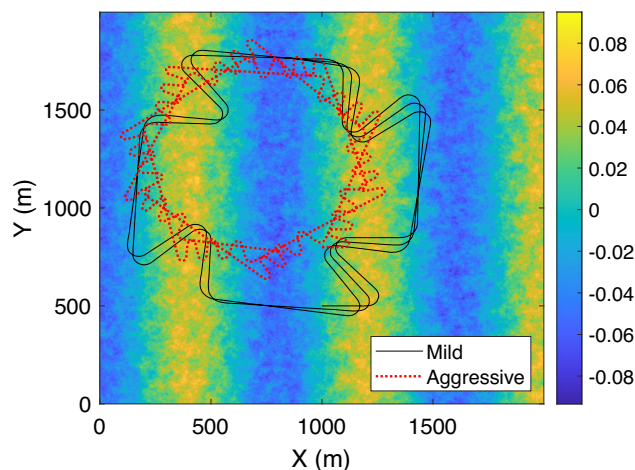


Fig. 8 Generated 2D road roughness profile based on ISO 8608-based spectral characteristics, with paths of mild and aggressive drivers

The resulting elevation map $Z(X, Y) \in \mathbb{R}^{N_X \times N_Y}$ spans a 2000×2000 m domain with 1 m spatial resolution. An example of the generated roughness is shown in Fig. 8, exhibiting spatial correlation with a standard deviation of approximately 0.045 m and peak-to-peak variation exceeding 0.15 m.

To simulate structured terrain features such as hills or ramps, we superimposed large-scale elevation patterns onto the stochastic roughness field. For instance, sinusoidal hills were defined in the longitudinal direction using:

$$Z_{\text{hill}}(X) = A_{\text{hill}} \cdot \sin \left(\frac{\pi(X - X_0)}{L_{\text{hill}}} \right)$$

where $A_{\text{hill}} = 0.05$ m is the hill amplitude, $X_0 = 1000$ m is the hill’s starting position, and $L_{\text{hill}} = 400$ m is the hill length. This design introduces realistic undulations similar to those found in rural or suburban roads.

To couple the road profile with vehicle dynamics, we implemented a spatial interpolator using `scipy.interpolate.RectBivariateSpline`, which provides smooth querying of $Z(X, Y)$ and its spatial gradients at arbitrary coordinates. The road elevation and partial derivatives $\partial Z/\partial X$ and $\partial Z/\partial Y$ were computed at each wheel based on the vehicle’s pose (position and yaw angle) and geometric configuration (wheelbase and track width). This allows direct integration of terrain excitation into the suspension dynamics model and the RL environment.

3.2.3 Computing wheel-level elevation and elevation rate

After generating the two-dimensional road profile $Z(X, Y)$ and defining the vehicle motion trajectories, we compute the road excitation input at each suspension point in the form of elevation \mathbf{z}_r and elevation rate $\dot{\mathbf{z}}_r$, which appear in Eq. (18)

of the suspension dynamics model. These quantities serve as exogenous inputs to the vertical dynamics of each wheel.

At each simulation timestep, the global position of the vehicle's center of gravity (CG), denoted as $(x(t), y(t))$, and its yaw angle $\psi(t)$, are used to compute the global positions of the four suspension contact points:

- Front left (FL): $(\Delta x, \Delta y) = (l_f, -l/2)$
- Front right (FR): $(\Delta x, \Delta y) = (l_f, l/2)$
- Rear left (RL): $(\Delta x, \Delta y) = (-l_f, -l/2)$
- Rear right (RR): $(\Delta x, \Delta y) = (-l_f, l/2)$

Each wheel's location is determined by rotating its local offset relative to the CG using the yaw angle:

$$\begin{aligned}x_{\text{wheel}} &= x + \Delta x \cos \psi - \Delta y \sin \psi, \\y_{\text{wheel}} &= y + \Delta x \sin \psi + \Delta y \cos \psi,\end{aligned}$$

where Δx and Δy are the relative longitudinal and lateral distances from the CG to each wheel, determined by l_f and l .

Given the global position of each wheel, the corresponding road elevation z_{ri} is extracted from the spatial map $Z(X, Y)$. In addition, the local terrain slopes $\partial Z/\partial X$ and $\partial Z/\partial Y$ are obtained from pre-computed gradient maps of the road surface. These values are queried using bilinear or spline-based interpolation to ensure smoothness.

The vertical rate of change of road elevation at each wheel contact point, denoted \dot{z}_{ri} , is calculated using the chain rule:

$$\dot{z}_{ri} = \frac{\partial Z_i}{\partial Y} v \cos \psi + \frac{\partial Z_i}{\partial X} v \sin \psi,$$

where v is the longitudinal velocity of the vehicle. This formulation allows the model to capture how the suspension experiences road gradients based on the direction and speed of vehicle motion.

3.3 Problem definition

The goal of active suspension control is to enhance both ride comfort and vehicle stability by actively adjusting the forces applied at each suspension. However, achieving optimal performance requires careful coordination between the physical design parameters (e.g., suspension stiffness and damping) and the control policy that governs the actuator behavior.

In our framework, we focus on co-designing the mechanical spring and damper parameters, i.e., k_s and c_s , for each of the four suspension units. These components play a critical role in determining how effectively the suspension system can isolate the vehicle body from road disturbances while maintaining contact between the tires and the ground.

To evaluate the performance of each suspension-controller configuration, we define a reward function that reflects key objectives in ride comfort, vehicle handling, and control efficiency. The reward is computed as a weighted sum of physically meaningful performance indices derived from the dynamic responses. Specifically, ride comfort is quantified using a comfort index, which aggregates vertical acceleration, pitch acceleration, and roll acceleration, defined by:

$$\text{comfort index} = \sqrt{(w_1 \ddot{z}_{sk})^2 + (w_2 \ddot{\alpha}_k)^2 + (w_3 \ddot{\beta}_k)^2}, \quad (20)$$

where $w_1 = 10$, $w_2 = 1.0$, and $w_3 = 0.5$ are scaling weights to account for their relative contributions to passenger discomfort, with vertical acceleration being the dominant factor.

In addition to comfort, handling performance is evaluated through penalties on pitch and roll angles, which are squared to reflect their increasing impact on stability as the magnitude grows. These penalties are normalized using empirically chosen scaling factors to ensure appropriate weighting within the total cost. Additionally, the reward also includes an energy penalty proportional to the squared control inputs applied at each suspension. Altogether, the reward function balances the competing objectives of comfort, handling, and control effort:

$$r_{k+1} = - \left(\text{comfort index} + c_1 \alpha_k^2 + c_2 \beta_k^2 + c_3 \sum_{i=1}^4 u_{ik}^2 \right), \quad (21)$$

where $c_1 = \frac{1}{0.00004}$, $c_2 = \frac{1}{0.00003}$, and $c_3 = 0.0001$ are coefficients according to the quantity scales.

The scale factors in Eqs. (20) and (21) are introduced to normalize the magnitudes of heterogeneous state variables and physical design parameters before they are provided as inputs to the neural networks. Their values are selected based on the typical operating ranges of the corresponding variables, ensuring that all inputs are of comparable numerical scale and improving training stability. These scale factors are fixed throughout training and are not tuned to optimize performance.

4 Multi-generation digital twin framework

Directly training RL agents on a physical system or a high-fidelity model is often impractical due to safety risks, computational cost, and potential system degradation during exploratory learning. Therefore, the proposed framework initializes learning in a low-fidelity digital model that provides a safe and efficient training environment. Data collected during nominal operation of the system are then used to update the digital model, enabling progressive improvement in fidelity

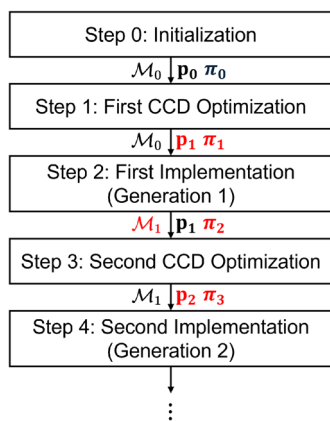


Fig. 9 Overview of the multi-generation digital twin-based control co-design framework. Updated components at each stage are highlighted in red. (\mathcal{M} : digital model, \mathbf{p} : system design, and π : controller)

without requiring risky online exploration on the physical system.

4.1 Overview

The framework proposed in (Tsai et al. 2025a) establishes a closed learning loop between the digital and physical domains, enabling continuous improvement of both system design and control policy across multiple generations. Building on this foundation, the present work applies and extends the framework to the optimization of a full-vehicle active suspension system, where the DT learns and adapts to distinct driving behaviors and varying environmental conditions. As illustrated in Fig. 9, the process begins with an initial policy π_0 and baseline digital model \mathcal{M}_0 (**Step 0**), followed by the first CCD optimization (**Step 1**) that jointly determines the physical design variables \mathbf{p}_1 and the control policy π_1 . These optimized configurations are then implemented on the physical platform (**Step 2**, Generation 1), where operational data are collected under realistic environmental and loading conditions.

During implementation, the DT acts as an intermediary that continuously synchronizes the digital and physical systems. Through real-time data acquisition and bi-directional communication, the digital model is updated to reflect the actual behavior of the physical system, producing an improved representation \mathcal{M}_1 . This learning process incorporates uncertainty quantification (UQ) and data-driven modeling to capture nonlinearities and environmental variations that were not represented in the initial model. At the same time, the control policy is refined based on the incoming data, leading to a self-evolving feedback structure that adapts to changing conditions. Prior to advancing to the next generation, the updated digital model \mathcal{M}_1 is used for the second CCD optimization (**Step 3**), in which both the physical design

and control policy are re-optimized. This iterative refinement process continues through successive generations, ensuring that the system design and controller become progressively more adaptive, efficient, and resilient to uncertainty.

Beyond enabling progressive model-policy co-evolution, the DT framework also supports customization of active suspension systems for different driving behaviors. As depicted in Fig. 10, driver behavior significantly influences suspension dynamics and performance objectives. For example, a mild driver typically generates smoother road excitations and moderate acceleration profiles, whereas an aggressive driver induces sharper transients and larger dynamic loads. After the first CCD optimization, the framework diverges into separate learning pathways for each driver type. Through physical data collection, real-time model and policy updates, and subsequent redesign with the updated models and policies, the framework yields tailored solutions for each driving profile. Specifically, the DTs evolve from $\mathcal{M}_0 \rightarrow \mathcal{M}_1$ for the mild driver and $\mathcal{M}_0 \rightarrow \mathcal{M}'_1$ for the aggressive driver. Re-optimization using these refined models produces the design-controller pairs (\mathbf{p}_2, π_3) and (\mathbf{p}'_2, π'_3) , which are customized to achieve optimal ride comfort, handling stability, and energy efficiency corresponding to each driving style.

Although the physical vehicle remains identical for both drivers, the resulting digital models \mathcal{M}_1 (mild) and \mathcal{M}'_1 (aggressive) differ because the quantile-based discrepancy model is updated using data collected under distinct driving behaviors, detailed in Sect. 4.4. Each driver generates unique dynamic excitations and statistical distributions of suspension responses. For instance, a mild driver generates smoother, low-frequency inputs with limited suspension travel, whereas an aggressive driver produces large-amplitude, high-frequency excitations and stronger coupling between lift, pitch, and roll motions. Consequently, the learned discrepancy functions calibrate the DTs within their respective operational regimes to generate behaviorally specialized models. These differences reflect context-dependent digital representations of the same physical system, enabling personalized DTs that accurately capture driver-specific dynamics and improve predictive fidelity for subsequent co-design and control optimization.

This multi-generation digital twin-enabled CCD framework therefore provides a unified methodology for continuous learning and adaptation to bridge model fidelity, real-time control, and design optimization to realize intelligent and personalized suspension systems for varying operational scenarios. It is important to note that the mild and aggressive driving cases are not intended to represent simultaneously deployable hardware configurations within a single vehicle. Instead, they serve as scenario-specific design studies that illustrate how optimal suspension design and control policies depend on assumed operating regimes. In practice, such results can inform system-level tuning and design deci-

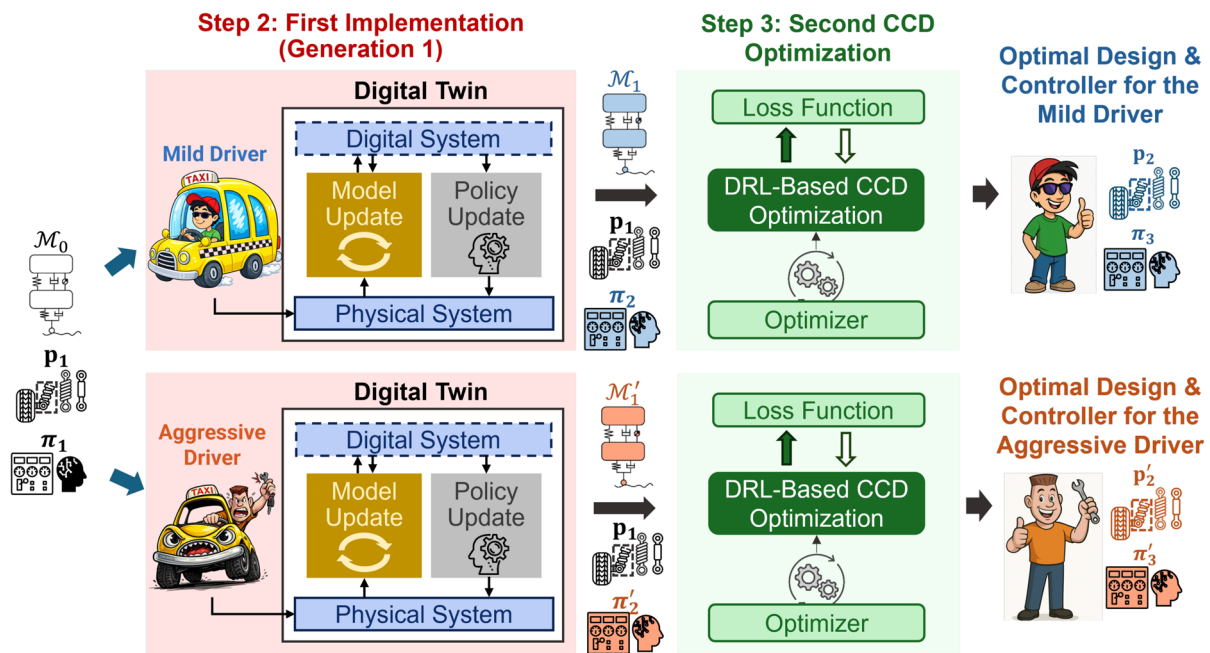


Fig. 10 Customization of active suspension designs and controllers for distinct driving behaviors after **Step 1** first CCD optimization. Through data collection, real-time model and policy updates, and redesign with the updated models and policies, we can yield customized and optimal solutions for different drivers. The mild and aggressive drivers exhibit distinct behaviors that significantly influence suspension dynam-

ics. By continuously updating the virtual systems (from $\mathcal{M}_0 \rightarrow \mathcal{M}_1$ and \mathcal{M}'_1) and control policies based on physical data, and re-optimizing the design-controller pair, the resulting solutions (\mathbf{p}_2, π_3) and (\mathbf{p}'_2, π'_3) are tailored to achieve optimal ride comfort and stability for each driver profile

sions or motivate the selection of a single robust suspension design that balances performance across heterogeneous driving behaviors.

4.2 Step 0: initialization

Prior to training the neural networks for the control policy, an initial set of control actions is generated to provide a stable starting point for the DRL optimization. This process serves as a warm start, in which state-action pairs derived from classical control laws are used to initialize the learning-based controller (Tsai and Malak 2022b). To compute approximate optimal actions for these sampled pairs under the nonlinear system dynamics, a set of proportional controllers (P controllers) is employed to regulate all state variables. The controller gains are tuned using Bayesian Optimization (BO) (Chen et al. 2024), which efficiently explores the continuous gain space while balancing exploration and exploitation.

The optimization problem is formulated to minimize the overall ride discomfort, quantified by the root mean square (RMS) value of the comfort index, defined in Eq. (20). The design vector consists of five controller parameters,

$$\mathbf{k} = [K_0, K_1, K_2, K_3, K_4], \tag{22}$$

which compactly represent the proportional gains for the observed states \mathbf{y} , including $\dot{z}_s, \dot{\alpha}, \dot{\beta}, z_{ui}$, and $z_s - z_{ui}$, for $i = 1, \dots, 4$. For each sampled gain vector $\mathbf{k} \in \mathbb{R}^{4 \times 11}$, defined as:

$$\mathbf{K} = \begin{bmatrix} K_0 & K_1 & K_2 & K_3 & 0 & 0 & 0 & K_4 & 0 & 0 & 0 \\ K_0 & -K_1 & K_2 & 0 & K_3 & 0 & 0 & 0 & K_4 & 0 & 0 \\ K_0 & K_1 & -K_2 & 0 & 0 & K_3 & 0 & 0 & 0 & K_4 & 0 \\ K_0 & -K_1 & -K_2 & 0 & 0 & 0 & K_3 & 0 & 0 & 0 & K_4 \end{bmatrix}. \tag{23}$$

where each row corresponds to one actuator input (u_1 to u_4) and each column corresponds to a specific state feedback term, the system is simulated under the closed-loop law $\mathbf{u} = -\mathbf{K}\mathbf{y}$, and the corresponding comfort index is evaluated. BO then updates its surrogate model to propose the next candidate \mathbf{k} that minimizes the comfort index.

After 120 iterations, the optimizer identifies the best set of gains $\mathbf{k}^* = [5000.0, 3000.0, 801.3, 10000.0, -1717.9]$ that minimizes the RMS comfort index, providing a physically reasonable and dynamically stable controller. This P-controller serves as the initial policy for the DRL-based CCD training, ensuring stable exploration and preventing divergence in the early learning stages.

We acknowledge that using the simplified controllers introduces approximation error and may yield suboptimal actions relative to the true nonlinear optimum. However, this approach remains valuable for high-dimensional systems. In particular, it provides a stable and informed initialization for the controller, helping to prevent instability and poor performance in the early stages of DRL training.

To initialize the DRL policy and value function, we employed fully connected feedforward neural networks with three hidden layers. The input to both networks is a 13-dimensional vector comprising 2 physical design parameters and 11 observed states. The policy network produces a 4-dimensional output corresponding to the control actions, while the value network outputs a single scalar value representing the state-value function. Each hidden layer in both networks contains 128 neurons with the hyperbolic tangent (Tanh) activation function applied after each layer to introduce nonlinearity and ensure smooth gradients. For the policy, two separate DNNs were used to model the mean and standard deviation of the action distribution. The samples along with the optimal actions $(\mathbf{x}_j, \mathbf{p}_j, \mathbf{u}_j^*)$ are used to initialize the mean DNN, whereas the weights and biases of the standard deviation DNN are initialized to zero and a small positive value (0.01) across all layers, respectively, to encourage initial exploration without introducing high variance.

4.3 Step 1: first CCD optimization

Following the initialization of the policy and value networks, the first stage of the multi-generation digital twin framework performs CCD of the full-vehicle active suspension system using a DRL algorithm based on PPO. The overall structure of the PPO-based CCD optimization process is illustrated in Fig. 11, which shows the interaction between the actor—critic networks, the environment, and the embedded physical design parameters.

Unlike standard PPO implementations that update only the neural-network parameters of the policy and value functions, the proposed DRL-based CCD formulation jointly optimizes both the physical suspension parameters (including the stiffness coefficient k_s and damping coefficient c_s of all four suspension systems) and the control policy parameters. This simultaneous optimization allows the physical system and its controller to co-evolve within the same learning loop, ensuring that the control policy adapts to hardware changes while the design variables are refined to improve closed-loop performance. It is noted that consistent with prior work in CCD, physical design parameters (in this case, $\mathbf{p} = [k_s, c_s]$) are modeled as continuous variables to enable gradient-based optimization and automatic differentiation, with discrete feasibility addressed through post-processing or component selection (Allison et al. 2014).

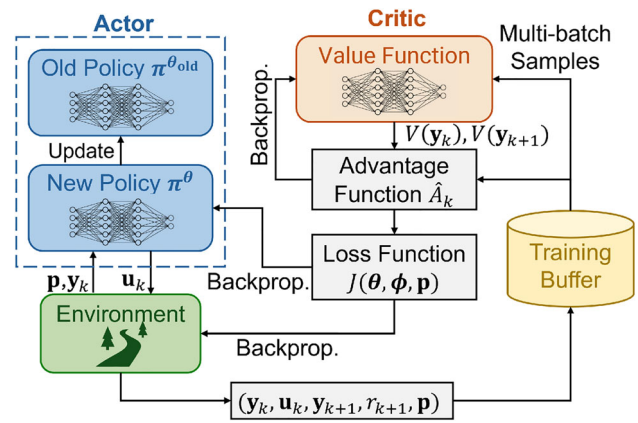


Fig. 11 Flow chart of the DRL-based CCD optimization using the PPO algorithm proposed by Tsai et al. (2025a). Unlike standard PPO implementations, the proposed framework embeds the physical design parameters \mathbf{p} into the input space, allowing simultaneous gradient-based updates of both the control policy and the physical design by treating the environment dynamics as differentiable and enabling backpropagation through the environment. The policy and value networks are updated using observed states \mathbf{y} since the full-vehicle suspension system is partially observable

The CCD problem minimizes a joint loss function composed of a clipped surrogate policy objective and the value loss:

$$\begin{aligned} & \min_{\theta, \phi, \mathbf{p}} J(\theta, \phi, \mathbf{p}) \\ & = \mathbb{E}_k \left[\underbrace{-\min \left(\rho_k(\theta, \mathbf{p}) \hat{A}_k, L^{\text{CLIP}}(\rho_k(\theta, \mathbf{p}), 1 - \epsilon, 1 + \epsilon) \hat{A}_k \right)}_{\text{Policy Loss}} \right. \\ & \quad \left. + c_v \cdot \underbrace{L_{\text{SmoothL1}} \left(V^\phi(\mathbf{x}_k, \mathbf{p}), \hat{V}_k \right)}_{\text{Value Loss}} \right], \end{aligned} \tag{24}$$

where \hat{A}_k is the advantage function at time step k , which estimates how much better an action is compared to the expected value of a state, and $L^{\text{CLIP}}(\rho_k(\theta, \mathbf{p}), 1 - \epsilon, 1 + \epsilon)$ is a function clipping $\rho_k(\theta, \mathbf{p})$ within $(1 - \epsilon, 1 + \epsilon)$, ρ_k denotes the probability ratio between the new and old policies:

$$\rho_k(\theta, \mathbf{p}) = \frac{\pi^\theta(\mathbf{u}_k | \mathbf{x}_k, \mathbf{p})}{\pi^{\theta_{\text{old}}}(\mathbf{u}_k | \mathbf{x}_k, \mathbf{p}_{\text{old}})}, \tag{25}$$

c_v is the coefficient for value loss, and we use Smooth L1 loss to define the value loss:

$$L_{\text{SmoothL1}}(a, b) = \begin{cases} \frac{1}{2}(a - b)^2, & \text{if } |x - y| < 1 \\ |a - b| - \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{26}$$

By embedding the suspension design vector \mathbf{p} (e.g., k_s and c_s for each suspension) into both policy and value net-

works, the optimization becomes differentiable with respect to the physical parameters, along with the parameters of the neural networks for the policy and value function. Gradients are computed through automatic differentiation using PyTorch autograd engine (Paszke et al. 2019), enabling end-to-end updates of all trainable variables. This mechanism effectively treats the entire digital twin model, including vehicle dynamics, controller, and design parameters, as a single computational graph, allowing the optimizer to explore both mechanical-design and control-policy spaces concurrently.

During training, the algorithm seeks to maximize ride comfort and handling stability by minimizing body acceleration, pitch/roll motion, and control effort. During training, the full-vehicle model is simulated under a constant forward velocity of 10 m/s, zero acceleration, and zero steering angle, representing a straight-line driving scenario. To emulate random road irregularities, the road disturbance $z_{r,i}$ and its time derivative $\dot{z}_{r,i}$ are sampled at each time step from zero-mean Gaussian distributions,

$$z_{r,i} \sim \mathcal{N}(0, 0.001^2), \quad \dot{z}_{r,i} \sim \mathcal{N}(0, 0.1^2), \quad (27)$$

applied independently to each of the four wheels ($i = 1, 2, 3$, and 4). The reward function penalizes body acceleration, pitch and roll rates, and control effort, driving the system toward an optimal trade-off between ride comfort, handling stability, and actuator energy efficiency.

Figure 12 illustrates the training history of the first co-design optimization. As shown in Fig. 12a, the average training return increases from approximately -2600 to -2063.6 over 2000 epochs, although with noticeable fluctuations due to the stochastic nature of RL. To more clearly assess performance improvement, we additionally report the evaluation return obtained using a deterministic policy with fixed physical parameters. This evaluation curve exhibits a clear upward trend, with the best performance achieved around epoch 1590, indicating that the CCD optimization effectively improves overall closed-loop system performance.

The policy learning rate is set to 1×10^{-5} , while the learning rate for plant (design) parameters is 5×10^{-5} . A discount factor of $\gamma = 0.99$ and generalized advantage estimation with $\lambda = 0.9$ are employed to balance bias and variance in long-horizon return estimation. Policy updates use a clipping parameter $\epsilon = 0.2$ with 20 optimization epochs per update. Each rollout consists of 20 time steps, with a buffer size of 20 and a minibatch size of 64. To ensure robust and stable co-design convergence, early stopping is determined based on evaluation performance rather than transient training peaks. Specifically, optimization is terminated when the relative improvement in evaluation return falls below 0.5% for 100 consecutive evaluation checkpoints and when the relative changes in stiffness and damping parameters remain

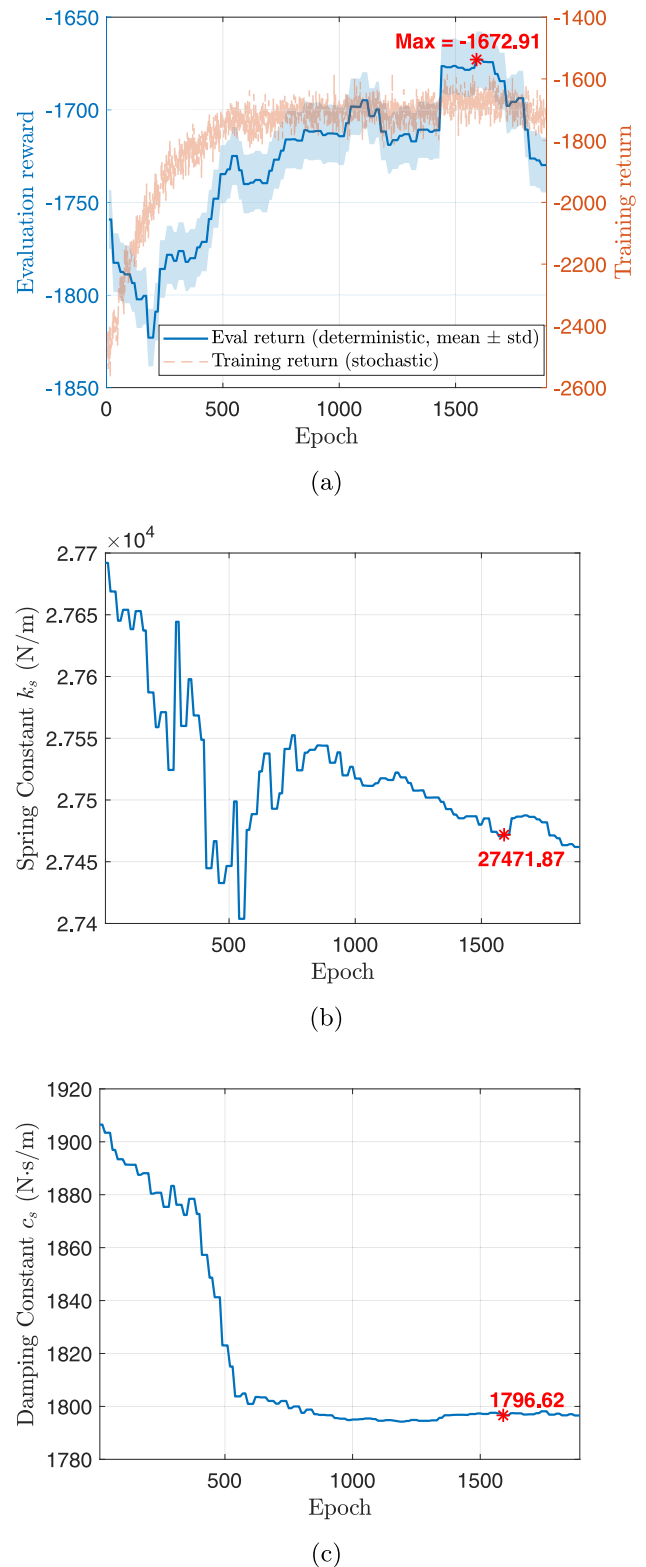


Fig. 12 Training history for the first CCD optimization with **a** average return, **b** system parameter: spring constant k_s , and **c** system parameter: damping constant c_s

below 0.2%. All evaluations are performed over a fixed horizon of 1000 time steps to provide a consistent basis for performance comparison.

The corresponding spring stiffness and damping coefficients associated with the best evaluation checkpoint are $k_s = 27471.87$ N/m and $c_s = 1796.62$ N·s/m, as shown in Fig. 12b and c, respectively. Rather than representing asymptotic convergence in a classical optimization sense, these values are selected based on a checkpoint-selection criterion that maximizes the evaluation return. This approach is standard in DRL, where the training objective is inherently stochastic and performance may degrade at later epochs due to exploration and distributional shift (Wexler et al. 2022; Eimer et al. 2023).

Although this simulation setup may appear simplified and not fully generalizable to diverse real-world scenarios, it represents a necessary first step before physical implementation, as it requires no experimental data or real-world measurements. In this first-stage study, the driving condition is intentionally simplified, assuming a constant vehicle speed (10 m/s) with Gaussian disturbances randomly added to the sprung-mass displacement (z_s) and velocity (\dot{z}_s). While this setup introduces controlled stochasticity, it also limits scenario diversity, making prolonged training susceptible to overfitting. To mitigate this issue, an early stopping criterion is employed, and the final design–policy pair ($\mathbf{p}_1, \boldsymbol{\pi}_1$) is selected as the checkpoint achieving the highest evaluation return. This configuration defines the Generation-1 design, which is deployed in the subsequent stage for physical implementation and data-driven model updating.

Compared with many existing studies, for example, (Dridi et al. 2025), which rely on deterministic bump or rectangular-road excitations, the proposed setup provides a more statistically representative training environment for early-stage co-design. This design enables the framework to capture fundamental stiffness–damping trade-offs and assess the stability of the DRL-based CCD algorithm before advancing to the next stage, where real-world road profiles and driving behaviors are incorporated through physical implementation within the digital twin loop.

Figure 13 presents representative time-domain trajectories comparing the PID and DRL controllers under identical disturbance realizations, whereas Table 2 summarizes the corresponding RMS performance metrics evaluated over multiple trajectories. During testing, the vehicle is driven at a nominal forward speed of 10 m/s with additive Gaussian noise ($\mathcal{N}(0, 0.5^2)$), small longitudinal acceleration perturbations ($\mathcal{N}(0, 0.2^2)$), and mild steering disturbances ($\mathcal{N}(0, 0.1^2)$), while random Gaussian road excitations are applied to each wheel through both road displacement and velocity inputs ($\mathcal{N}(0, 0.001^2)$ and $\mathcal{N}(0, 0.1^2)$, respectively). Under these conditions, the DRL controller consistently suppresses pitch and roll motions more effectively than the

PID controller, as evidenced by significant reductions in RMS pitch angle, roll angle, and their corresponding velocities. These improvements translate directly to enhanced ride stability and handling performance. Moreover, the DRL controller achieves a reduction in body vertical acceleration, indicating improved ride comfort under stochastic road excitations. Although the DRL controller requires slightly higher average control effort, the resulting closed-loop performance gains demonstrate a favorable trade-off between comfort, stability, and actuation demand.

Importantly, these results clarify the added value of the proposed RL-based CCD framework relative to classical optimization and model-based control strategies. The performance gains observed in Fig. 13 and Table 2 arise from the joint optimization of physical suspension parameters and control policy based on closed-loop performance under stochastic operating conditions, rather than from controller tuning alone. Unlike conventional plant optimization methods that rely on deterministic or simplified disturbance assumptions, the RL-based formulation evaluates candidate plant designs through trajectory-level interaction with the controller under randomized speed perturbations, steering noise, and multi-wheel road excitations. Furthermore, applying a traditional PID controller to the optimized hardware parameters confirms that the improvement stems from the integrated co-design of plant and control, highlighting the advantage of RL-based co-design in capturing nonlinear dynamics, uncertainty, and plant–control coupling.

4.4 Step 2: first implementation (generation 1)

Before deploying the optimized design and policy from Step 1 to the real environment, it is essential to recognize and quantify the mismatch between the digital model and the physical vehicle system. Even with high-fidelity modeling, discrepancies inevitably arise from nonlinearities not being captured, component variability, and environmental factors not represented in the nominal digital model. In practice, these mismatches lead to performance degradation if the control policy is transferred directly without adaptation.

To emulate such discrepancies, we construct a hypothetical real system that introduces realistic deviations from the nominal simulation model in three primary ways:

1. **Nonlinear spring and damper behavior.** In real suspensions, stiffness and damping characteristics are rarely linear across the full stroke and velocity range. To account for this, the spring and damper forces are augmented with cubic and velocity-dependent nonlinear terms:

$$F_{S_i} = k_s(z_{u,i} - z_s - \Delta_i) + c_s(\dot{z}_{u,i} - \dot{z}_s - \dot{\Delta}_i) + k_{nl}(z_{u,i} - z_s - \Delta_i)^3$$

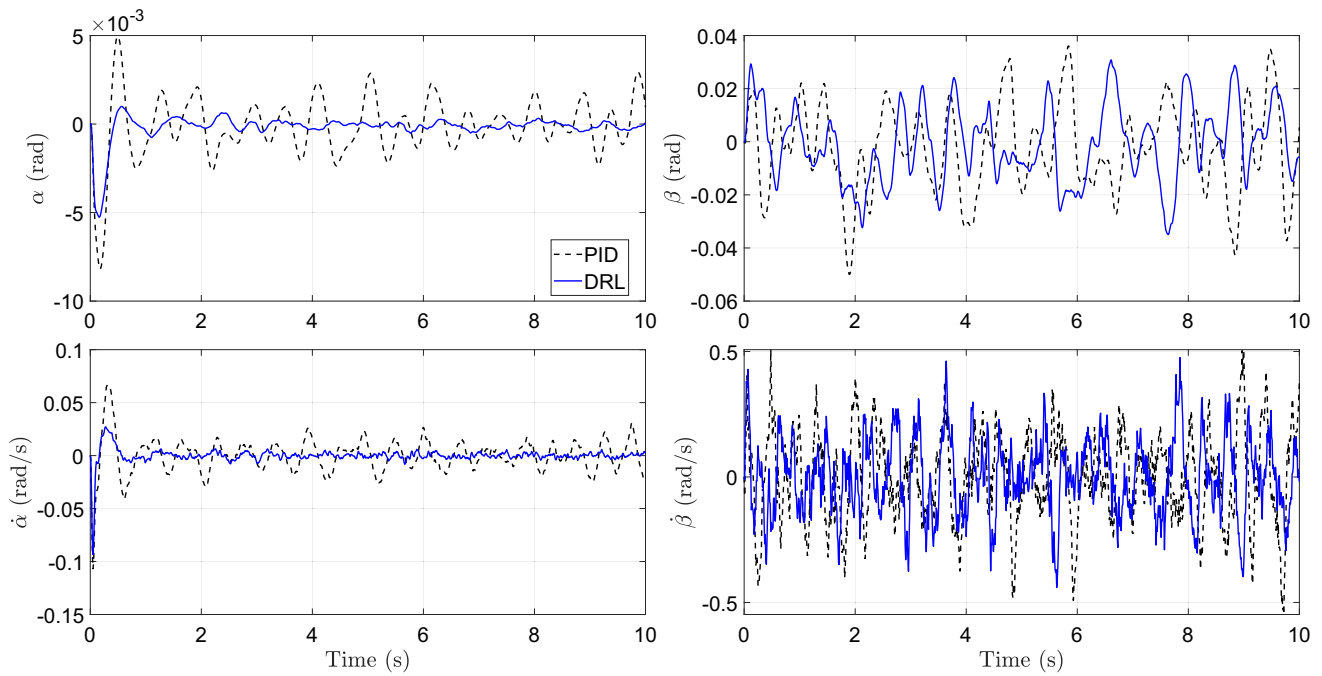


Fig. 13 Comparison of trajectories by PID and DRL controllers

Table 2 Comparison of performance metrics between PID and RL controllers

Metric	PID	RL
RMS pitch angle α (rad)	1.736×10^{-3}	7.720×10^{-4}
RMS roll angle β (rad)	1.696×10^{-2}	1.426×10^{-2}
RMS pitch velocity $\dot{\alpha}$ (rad/s)	1.666×10^{-2}	7.263×10^{-3}
RMS roll velocity $\dot{\beta}$ (rad/s)	1.802×10^{-1}	1.536×10^{-1}
RMS body vertical acceleration \ddot{z}_s (m/s ²)	2.673×10^{-1}	2.595×10^{-1}
Control effort mean of $ \mathbf{u}_k $ (N)	47.529	50.350
Return $\sum_{k=0}^{1000} r_{k+1}$	-14514.428	-11686.332

$$+c_{nl}|\dot{z}_{u,i} - \dot{z}_s - \dot{\Delta}_i|(\dot{z}_{u,i} - \dot{z}_s - \dot{\Delta}_i), \tag{28}$$

where $k_{nl} = 0.1k_s$ and $c_{nl} = 0.1c_s$. This nonlinear formulation captures the amplitude-dependent stiffness and rate-sensitive damping observed in real suspension systems.

2. **Non-uniform unsprung masses.** Manufacturing variability, tire-wheel assemblies, and sensor packages introduce asymmetric wheel masses. The unsprung mass vector is therefore defined as

$$\mathbf{m}_u = [60.0, 50.0, 45.0, 50.0] \text{ kg}, \tag{29}$$

representing the front-left, front-right, rear-left, and rear-right wheels, respectively.

3. **Asymmetric vehicle geometry.** To reflect potential variations in the structural layout and load distribution, the center-of-gravity height is perturbed by +0.05 m relative

to the nominal model, giving

$$h_{cg} = 0.55 + 0.05 \text{ m}. \tag{30}$$

This asymmetry slightly shifts the pitch and roll dynamics, leading to differential responses across the front and rear suspensions.

4. **Heterogeneous tire stiffnesses across wheels.** Real tires exhibit corner-dependent stiffness due to wear, load, and construction differences. We model this by assigning wheel-specific tire stiffnesses

$$\mathbf{k}_t = [0.9, 1.2, 1.1, 0.9] \times 200,000 \text{ N/m}, \tag{31}$$

for the front-left, front-right, rear-left, and rear-right wheels, respectively. The tire forces are then

$$F_{T_i} = k_{t,i}(z_{r,i} - z_{u,i}) + c_t(\dot{z}_{r,i} - \dot{z}_{u,i}). \tag{32}$$

5. **Deviations in sprung mass and moments of inertia.** To represent loading variation and structural uncertainty, the sprung mass and rotational inertias are increased by 10% relative to the nominal values:

$$m_s = 1.1 \times 1500 = 1650 \text{ kg}, \tag{33}$$

$$I_\theta = 1.1 \times 2500 = 2750 \text{ kg} \cdot \text{m}^2, \tag{34}$$

$$I_\phi = 1.1 \times 500 = 550 \text{ kg} \cdot \text{m}^2. \tag{35}$$

These deviations alter the heave–pitch–roll couplings and shift natural frequencies, increasing the reality gap for controller transfer.

Together, these modifications define the Generation 1 physical system, which serves as the “real” environment in the digital twin framework. The policy π_1 and design parameters \mathbf{p}_1 obtained from **Step 1** are directly deployed in this perturbed environment to evaluate transfer performance under realistic uncertainty.

Despite the discrepancies between the digital model and the physical system, the DRL-based controller demonstrates strong robustness and adaptability. This resilience arises from the inherently adaptive nature of RL, which enables the policy to evolve in response to changes in system dynamics and environmental conditions. Within the digital twin framework, sensor data collected from the physical vehicle (e.g., suspension deflections, actuator forces, and body accelerations) are continuously streamed to the digital model. These real-time observations are used to update the policy (π_2). This allows the agent to refine its decision-making strategy based on the latest system behavior. Through this process, the RL agent learns to maintain closed-loop stability and optimal ride–handling balance even under uncertainty by gradually compensating for model–reality discrepancies, nonlinear effects, and asymmetric parameter variations introduced in the physical system.

To explicitly account for the mismatch between the digital and physical systems, a discrepancy model is developed using physical data collected during the Generation 1 deployment. This model quantifies deviations between predicted and observed system responses by leveraging quantile regression (Chen et al. 2025b), a data-driven uncertainty quantification (UQ) technique that estimates conditional quantiles of a response variable. Unlike traditional regression methods, quantile regression provides predictive intervals that mainly capture the uncertainties from biased parameters and environmental noises (Chen et al. 2025b). Moreover, this approach does not need to assume Gaussian noise, thereby improving its flexibility in representing nonlinear, asymmetric errors commonly observed in real vehicle dynamics. Different from online calibrating model parameters (Chen and Chan 2021), we argue that a discrepancy model directly

applied to correct the system dynamic model allows better generalization while capturing missing physics.

In practice, the DT continuously collects real-time data from onboard sensors to monitor the vehicle state as well as driving conditions such as forward acceleration, vehicle speed, and steering angle. These measurements provide the essential information required to identify and update the discrepancy between the digital model and the real system. Through learning from this data stream, the discrepancy model dynamically refines its estimates of modeling error during operation. Importantly, the data used for model updating are assumed to be collected during normal system operation, rather than through direct RL interaction with the physical system.

By training the model at representative quantile levels, typically the 10th, 50th (median), and 90th percentiles, the DT captures both the central tendency and spread of the discrepancy distribution. The discrepancy at the next time step is estimated as a function of the current observed states, control actions, previous errors, and the driving conditions (forward acceleration a , speed v , and steering angle δ):

$$\begin{bmatrix} \mathbf{e}_{k+1}^{\text{upper}} \\ \mathbf{e}_{k+1}^{\text{median}} \\ \mathbf{e}_{k+1}^{\text{lower}} \end{bmatrix} = \mathbf{f}_e(\mathbf{e}_k, \mathbf{y}_k, \mathbf{u}_k, a_k, v_k, \delta_k), \tag{36}$$

where $\mathbf{e}_k := \mathbf{y}_k - \hat{\mathbf{y}}_k$ denotes the deviation between the actual state \mathbf{y}_k and the nominal predicted state $\hat{\mathbf{y}}_k$ from the digital model. The predicted quantiles $\mathbf{e}_{k+1}^{\text{upper}}$, $\mathbf{e}_{k+1}^{\text{median}}$, and $\mathbf{e}_{k+1}^{\text{lower}}$ represent the range of the deviations at the next step.

As shown in Fig. 14, the learned quantiles effectively bound the observed trajectories of the first three system states and capture both the variability and the bias between the digital model predictions and the real physical responses. The root mean square errors for the validation dataset are 0.0147 and 0.0151 under mild and aggressive driving profiles, respectively. The median quantile tracks the mean trajectory, while the upper and lower quantiles define the confidence region representing the dynamic uncertainty envelope of the physical system. This learned discrepancy model is then integrated into the DT to adaptively calibrate predictions and guide the controller update for subsequent generations.

While the learned quantiles successfully capture most of the real trajectories, a few observed data points fall outside the predicted uncertainty bounds, particularly in the third state corresponding to the roll velocity $\dot{\beta}$ (see Fig. 14). This behavior stems from an unbalanced dataset, where the frequency of extreme operating conditions is considerably lower than that of regular driving states. As a result, the quantile regression model tends to underestimate the probability of rare events and overestimate uncertainty during nominal conditions. Nevertheless, as more diverse data become available

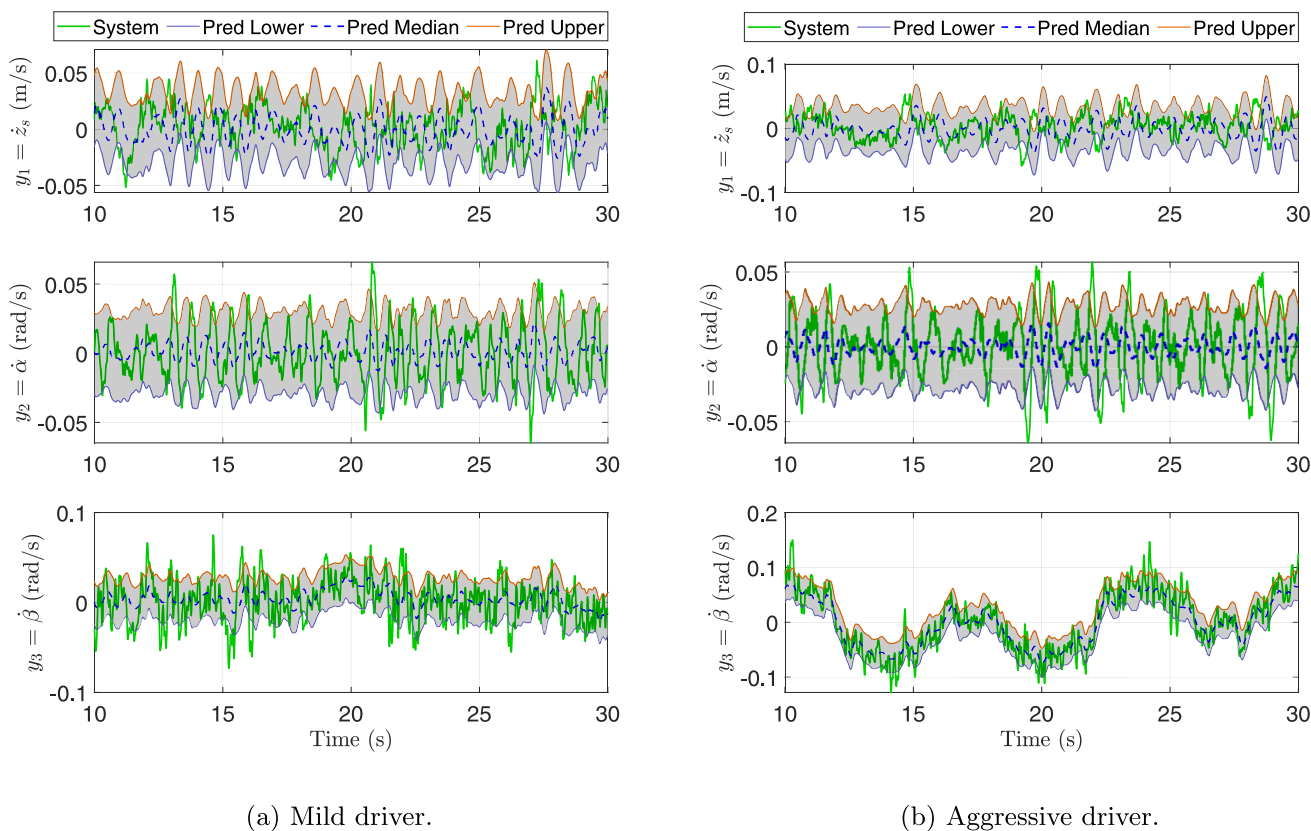


Fig. 14 Visualization of the learned quantiles and the real system trajectories of the first three observed states for the active suspension system

through subsequent generations of physical operation, the quantile estimates are expected to become more accurate and better calibrated across the full operational spectrum.

4.5 Step 3: second CCD optimization

The primary distinction between the first and second CCD optimizations lies in the integration of the updated digital twin model that incorporates the quantile-based discrepancy learning derived from Eq. (36). With the physical data collected during the first implementation, the digital model now provides a more accurate and uncertainty-aware representation of the vehicle dynamics. Unlike the initial optimization from Step 1, which relies on stochastically generated Gaussian disturbances to emulate environmental variability, the second CCD optimization directly leverages the learned discrepancy model to reflect realistic uncertainty in both the dynamics and external excitations. The updated model is constructed by incorporating the learned discrepancy function from Eq. (36) into Eq. (18), defined as:

$$\begin{bmatrix} \mathbf{y}_{k+1}^{\text{upper}} \\ \mathbf{y}_{k+1}^{\text{median}} \\ \mathbf{y}_{k+1}^{\text{lower}} \end{bmatrix} = \mathbf{C}\mathbf{x}_{k+1}$$

$$+\mathbf{f}_{\mathbf{e}}(\mathbf{e}_k, \mathbf{y}_k, \mathbf{u}_k, a_k, v_k, \delta_k) \tag{37}$$

where \mathbf{x}_{k+1} is the next state estimated by the model from Eq. (18). The optimization process no longer depends on artificially injected noise; instead, it inherently embeds the variability observed in the real system through the model. This advancement enables a more faithful simulation environment for DRL, leading to improved transferability and faster policy convergence in subsequent generations.

In the second CCD optimization, the new reward function is extended from Eq. (21) to explicitly account for the uncertainties captured by the quantile-informed digital model, leading to:

$$r'_{k+1} = -(r_{k+1} + \lambda_u J_{\text{uncertainty}}), \tag{38}$$

where

$$J_{\text{uncertainty}} = J_{\text{comfort,unc}} + c_1 J_{\text{pitch,unc}} + c_2 J_{\text{roll,unc}}. \tag{39}$$

The corresponding uncertainty term is computed as

$$J_{\text{comfort,unc}} = \sqrt{(w_1 \Delta a_{z,k})^2 + (w_2 \Delta \ddot{\theta}_k)^2 + (w_3 \Delta \ddot{\phi}_k)^2}, \tag{40}$$

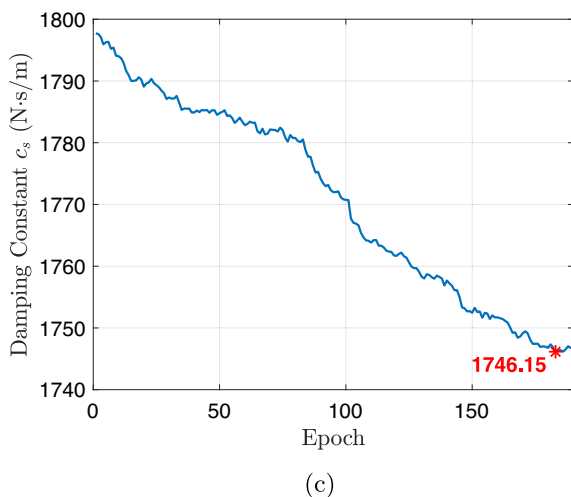
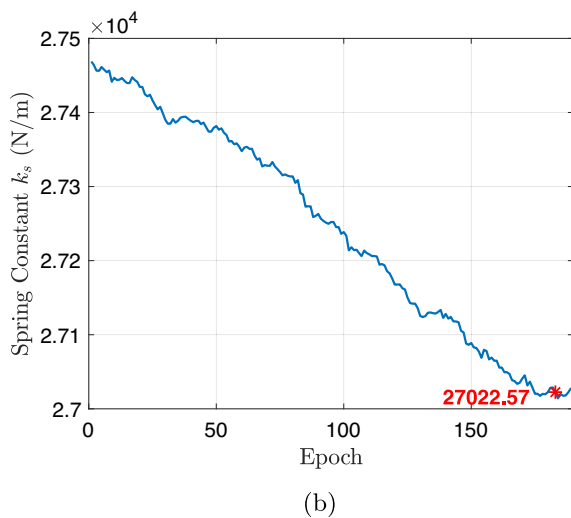
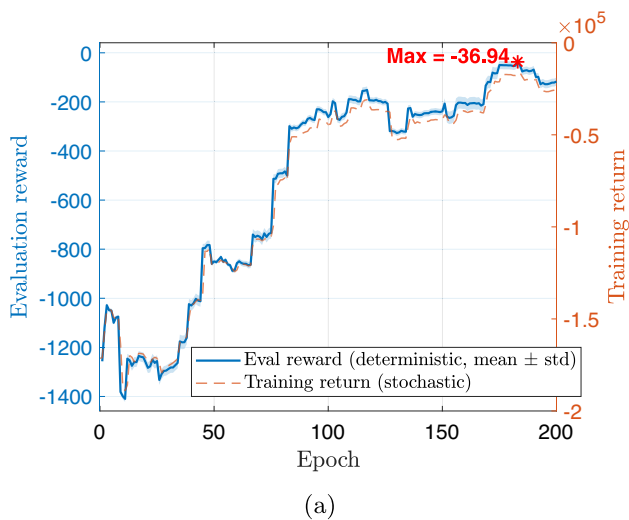


Fig. 15 Training history of the second CCD optimization for the mild driver, with **a** return, **b** system parameter: spring constant k_s , and **c** system parameter: damping constant c_s

where $\Delta(\cdot)$ represents the prediction uncertainty estimated by the quantile regression model. Similarly, $J_{\text{pitch,unc}} = (\Delta\alpha_k)^2$ and $J_{\text{roll,unc}} = (\Delta\beta_k)^2$ represent their corresponding uncertainty penalties. This modified formulation encourages the agent to not only optimize nominal performance but also achieve robust behavior by minimizing the propagation of model and environmental uncertainties.

Figure 15 shows the training histories of the return, spring stiffness k_s , and damping coefficient c_s for the mild driver over approximately 200 epochs. A new best model was identified at epoch 183 with a reward of -36.94 . The training follows the same PPO-based architecture and hyperparameter settings as those used in the first CCD optimization (Sect. 4.3). The key difference lies in the treatment of the system dynamics horizon and convergence criteria, reflecting the increased fidelity of the updated digital twin and the longer driving scenarios considered. Specifically, the total simulation horizon is extended to 120,000 time steps to capture the full driving profile, and the rollout length is increased to 100 time steps to ensure stable policy gradient estimation over longer temporal correlations. Due to the substantially higher computational cost associated with the extended horizon, the number of training epochs is reduced accordingly. Early stopping is applied using a more stringent window of 5 consecutive evaluation checkpoints, while the relative improvement threshold and design stabilization criteria remain consistent with those of the first CCD optimization. These settings enable efficient yet stable co-design refinement under long-horizon dynamic conditions.

It is observed that the training return exhibits a brief but pronounced drop around epochs 190–210 before recovering to its nominal range (Fig. 15a). Such transient degradations are a known characteristic of DRL-based co-design problems and arise from the combined effects of stochastic policy exploration and environmental nonstationarity. During training, the PPO agent samples actions from a stochastic policy to explore both the control and design spaces, which can occasionally produce rollouts associated with unusually poor performance. Because policy updates rely on on-policy rollouts, these exploratory trajectories may momentarily dominate the training return even when the overall learning trend remains favorable.

In addition, unlike control-only optimization, the second CCD stage jointly updates the physical design variables (k_s, c_s) together with the control policy. These design updates temporarily alter the system dynamics, introducing nonstationarity into the learning environment. When a co-design update shifts the system toward a less favorable region of the design space (such as an overly compliant or insufficiently damped suspension configuration), the resulting mismatch between the current policy and the updated dynamics can lead to a sharp but temporary reduction in return. This mechanism explains the localized return collapse observed near

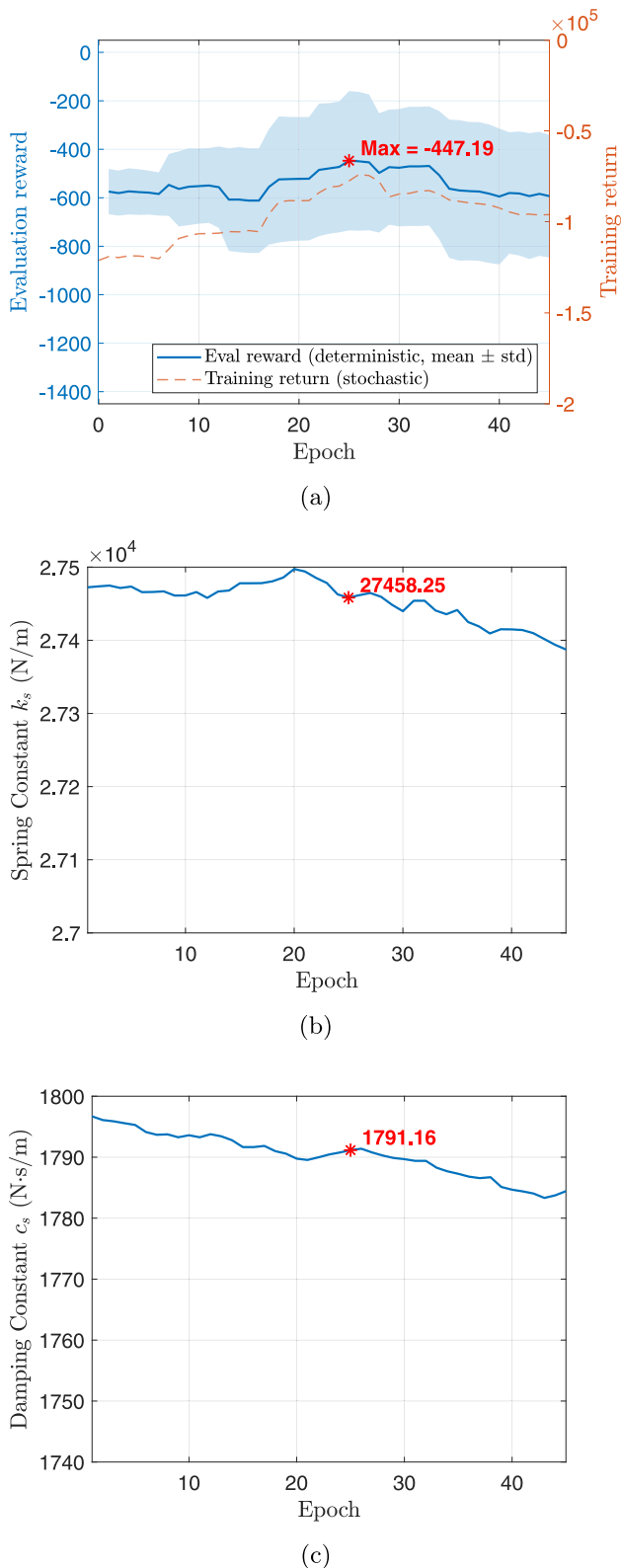


Fig. 16 Training history of the second CCD optimization for the aggressive driver, with **a** return, **b** system parameter: spring constant k_s , and **c** system parameter: damping constant c_s

epoch 200, which coincides with noticeable adjustments in the design parameters shown in Fig. 15b and Fig. 15c.

Importantly, despite these transient effects in the training return, the deterministic evaluation return exhibits a smooth and monotonic improvement trend, ultimately reaching a stable plateau beyond epoch 142. The clipped surrogate objective of PPO, together with conservative learning rates and trust-region-style co-design updates, prevents these disturbances from destabilizing the learning process. As a result, the optimization converges to a robust and stable solution, with both the physical design parameters and closed-loop performance remaining consistent once the evaluation return has plateaued.

Figure 16 presents the training histories for the aggressive-driving scenario. Compared with the mild-driver case, the second CCD optimization converges much earlier, with the best model identified at epoch 25 and an average reward of -447.19 . Afterward, no further improvement was observed, and the training quickly reached the patience threshold for early stopping. The faster convergence arises because the updated digital model \mathcal{M}'_1 effectively captured the more energetic driving conditions. Another contributing factor is that, under aggressive driving, the system operates closer to its physical and control limits, leaving less room for further improvement through policy or design adjustments. As a result, the learning process quickly stabilizes around a near-optimal solution that balances ride comfort, stability, and control effort within the achievable performance envelope of the suspension system.

The optimized design parameters converged to $k_s = 27,458.25$ N/m and $c_s = 1791.16$ N·s/m, both higher than those of the mild-driver design, which aligns with the physical intuition that stiffer and more strongly damped suspensions are required to maintain stability under aggressive driving maneuvers. These results further confirm that the multi-generation digital twin framework can efficiently adapt and customize the suspension design and control policy for distinct driving behaviors.

5 Results and discussion

5.1 Suspension designs

Table 3 summarizes the spring and damping constants (k_s , c_s) for different stages of the CCD process (initial, optimized after first CCD, and optimized after second CCD for two driving scenarios). These parameters collectively define the passive dynamics of the suspension, which interact with the active control system to achieve desired ride and handling performance. The relationship between the physical suspension design and the control system can be interpreted through the framework of open-loop and closed-loop co-

Table 3 Summary of initial and optimized suspension design parameters (k_s : spring constant and c_s : damping constant)

	k_s (N/m)	c_s (N·s/m)
Initial	27,692.00	1906.50
Optimized after first CCD	27,471.87	1796.62
Optimized after second CCD (mild)	27,022.57	1746.15
Optimized after second CCD (aggressive)	27,458.25	1791.16

Table 4 Performance comparison before and after the second CCD optimization for mild and aggressive drivers

	Metric	Before	After	Improvement
Mild	RMS of body acceleration (m/s^2)	0.2648	0.2189	17.32
	Control effort (Mean of $ \mathbf{u}_k $, N)	214.897	147.541	58.82
Aggressive	RMS of body acceleration (m/s^2)	0.3705	0.2664	28.08
	Control effort (Mean of $ \mathbf{u}_k $, N)	228.978	200.402	12.48

Improvements are computed as $|\text{Before} - \text{After}|/\text{Before} \times 100\%$

design (Deshmukh et al. 2015). A softer suspension (lower k_s and c_s) implies that the passive system alone provides less vibration suppression and stability; however, this allows the active controllers to play a larger role in compensating external disturbances and maintaining ride comfort. Conversely, a stiffer suspension limits the range of active control action but enhances the inherent stability and responsiveness of the vehicle.

For the mild-driver scenario, the second CCD optimization results in significantly reduced stiffness and damping ($k_s = 27022.57$ N/m, $c_s = 1746.15$ N·s/m). This softer configuration indicates that the control system effectively assumes a greater share of the vibration mitigation task, actively stabilizing the vehicle and isolating the body motion from road irregularities. Such a design is physically intuitive, as mild driving introduces smaller disturbances, allowing the actuators to manage vibrations without large control effort while maintaining comfort.

In contrast, the aggressive-driving scenario exhibits a stiffer configuration ($k_s = 27458.25$ N/m, $c_s = 1791.16$ N·s/m), reflecting the higher demand for dynamic stability and fast response under harsh maneuvers. The system favors a more rigid suspension that provides stronger passive support to resist large load transfers and rapid body oscillations, thereby reducing the burden on the controller. This trend aligns with the intuition from the active-passive co-design literature (Deshmukh et al. 2015), where aggressive operating conditions drive the optimization toward increased structural stiffness and damping to maintain controllability and robustness. Overall, these results reveal how the CCD framework systematically balances the contribution of passive and active elements, leading to physically interpretable suspension designs that adapt to different driving behaviors.

5.2 Performance comparison

Table 4 compares the system performance before and after the second CCD optimization. Two performance metrics are evaluated: (1) ride comfort, represented by the root mean square (RMS) of body vertical acceleration, and (2) control effort, computed as the mean of the absolute values of the four actuator forces. Importantly, these performance metrics are evaluated under perturbed driving and road conditions that differ from the profiles used during the optimization stage. Specifically, stochastic perturbations are applied to the driving inputs and road excitations: the vehicle forward speed is corrupted with zero-mean Gaussian noise with a standard deviation of 0.25 m/s, the longitudinal acceleration with Gaussian noise of 0.05 m/s^2 , and the steering angle with Gaussian noise of 0.1°. In addition, random disturbances are applied to the wheel-level road elevation and elevation rate signals. This ensures that the evaluation reflects an unseen test scenario rather than the optimization targets themselves.

Although the RMS acceleration decreases only slightly for the mild driver (about 17%), the control effort is reduced by more than 50%. This outcome suggests that the second CCD optimization enhances control efficiency by refining the coordination between the suspension hardware and the control policy, achieving comparable comfort with substantially less actuation energy. The persistence of this improvement under stochastic perturbations indicates that the learned design-policy pair generalizes beyond the nominal mild-driving profile used during optimization.

For the aggressive driver, the RMS body acceleration is reduced by approximately 28%, indicating a substantial improvement in ride comfort under high-intensity driving conditions. At the same time, the control effort decreases by about 12%, demonstrating that the second CCD optimization achieves improved vibration suppression without relying on excessive actuation. These results suggest that

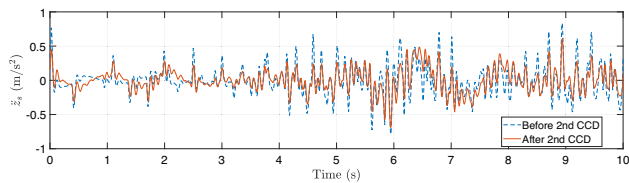


Fig. 17 Trajectories of the vertical body acceleration for the mild-driver scenario before and after the second CCD optimization. Only the first 10 s of the trajectories are shown for clarity

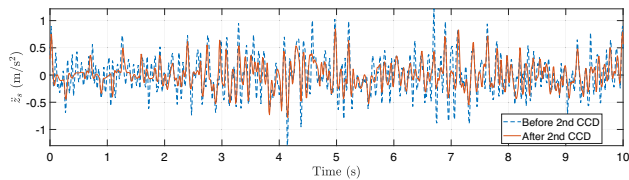


Fig. 18 Trajectories of the vertical body acceleration for the aggressive-driver scenario before and after the second CCD optimization. Only the first 10 s of the trajectories are shown for clarity

the co-designed suspension and control policy more effectively manage aggressive road excitations by leveraging the underlying suspension dynamics while maintaining efficient control actions. Notably, these gains are observed despite the presence of randomized variations in driving inputs and road disturbances, demonstrating the robustness of the co-designed suspension and control policy to previously unseen operating conditions. Overall, these results highlight the ability of the CCD framework to yield energy-efficient control without compromising ride quality.

5.3 Analysis of trajectories

Figures 17 and 18 compare the trajectories of the vertical body acceleration \ddot{z}_s (which is one of the most critical states for riding comfort) before and after the second CCD optimization for the mild- and aggressive-driving scenarios, respectively. In both cases, the trajectories after the second CCD exhibit consistently reduced magnitudes, with the most pronounced improvement observed in the peak accelerations. Because peak vertical accelerations are closely associated with perceived ride discomfort and fatigue, their attenuation indicates a meaningful improvement in ride quality. Notably, these reductions are achieved across both driving scenarios, demonstrating that the second CCD stage enhances closed-loop performance in a manner that is robust to changes in driving style and disturbance intensity.

Beyond peak reduction, the post-optimization trajectories appear smoother, with fewer abrupt fluctuations and more uniform responses over time. This behavior suggests improved coordination between the suspension design parameters and the control policy, enabling the system to better dissipate road-induced disturbances and mitigate transient

shocks. Importantly, the improvements are achieved without introducing high-frequency oscillations or excessive control activity, indicating a balanced trade-off between comfort and control effort. Together, these results confirm that the second CCD optimization refines both the physical–control interaction and the system’s dynamic response, leading to improved ride comfort and stability under diverse operating conditions.

Figures 19 and 20 illustrate the actuator force trajectories (u_1 – u_4) before and after the second CCD optimization for mild and aggressive drivers, respectively. The dashed blue lines denote the actuator forces before optimization, and the solid orange lines represent those after optimization. Only the first 120 s of trajectories are shown for clarity.

For the mild driver, the first actuator exhibits a pronounced reduction in force magnitude after the second CCD step, and the third and fourth actuators (rear suspensions) also show large decreases. These reductions and adjustments reflect the controllers’ ability to adaptively redistribute control effort in response to the propagated body motion induced by front-wheel irregularities. By sensing the resulting pitch and lift dynamics through the available measurements, the rear suspension controllers *proactively* coordinate their actions to stabilize the vehicle body, illustrating the closed-loop adaptability of the learning-based control policy. Also, these changes indicate smoother and more coordinated control actions, contributing to improved comfort and reduced energy consumption.

For the aggressive driver, while the system still demands greater overall actuation to handle harsher maneuvers, most actuator trajectories become smoother with fewer abrupt peaks. The optimization thus preserves the responsiveness needed for aggressive driving while achieving better coordination among actuators.

Overall, the second CCD step significantly reduces control magnitudes and smooths actuator responses for both driver types. These results demonstrate the framework’s capacity to adaptively tune physical and control designs to distinct driving behaviors.

6 Conclusion

This work presents a multi-generation digital twin (DT)-based control co-design (CCD) framework for full-vehicle active suspension system to optimize the ride comfort and minimize the energy consumption. This framework integrates deep reinforcement learning (DRL), specifically the Proximal Policy Optimization (PPO) algorithm to jointly optimize the physical suspension components (spring stiffness and damping coefficient) and active controllers. By combining DTs and real-time updating the uncertainty-aware models, the framework enables the systems to learn and evolve with operational data, driver behaviors, and environ-

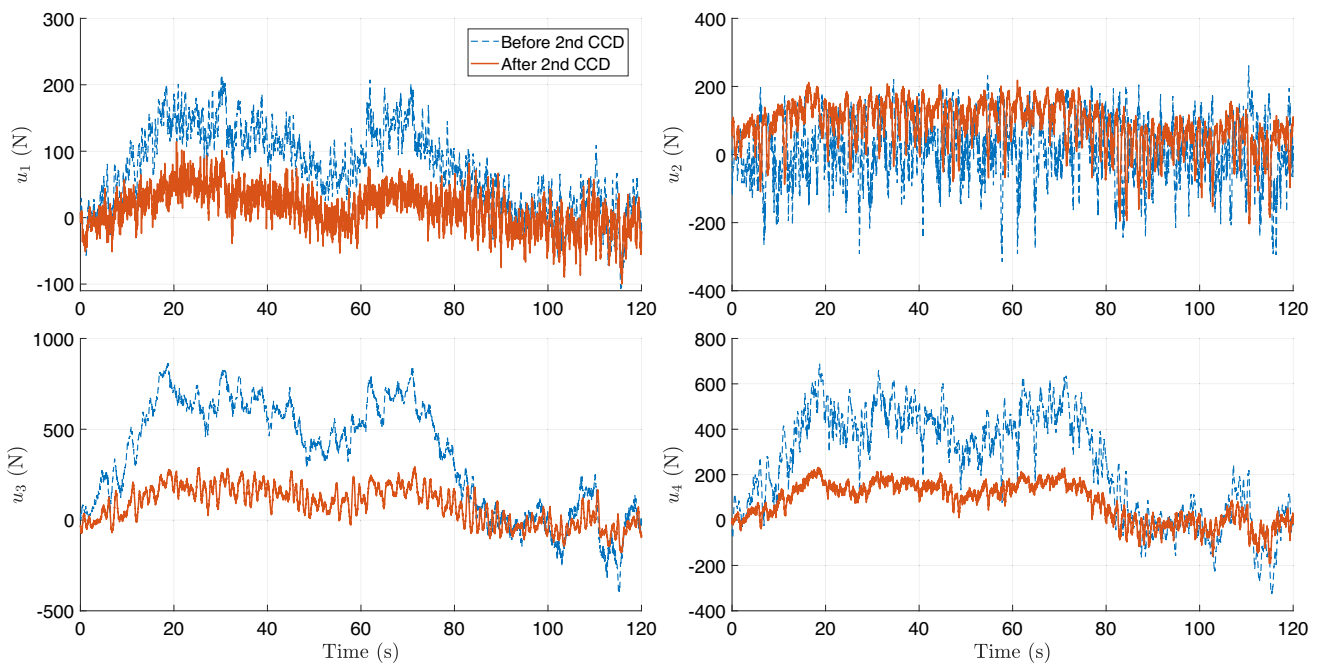


Fig. 19 Control trajectories of the mild-driver scenario before and after the second CCD optimization. Only the first 120s of the trajectories are shown for clarity

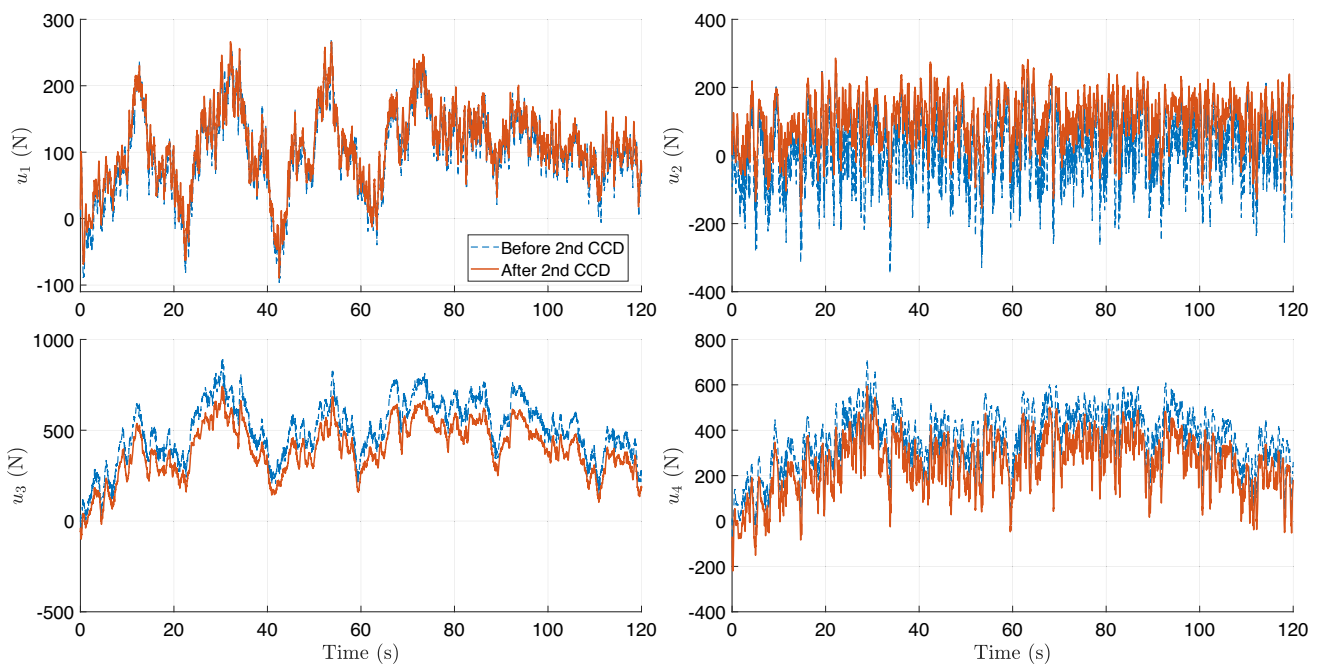


Fig. 20 Control trajectories of the aggressive-driver scenario before and after the second CCD optimization. Only the first 120s of the trajectories are shown for clarity

mental conditions. With two distinct driving scenarios (mild and aggressive drivers), the case study demonstrates that the method can perform personalized optimization by tailoring both the physical parameters and control policies for different conditions and requirements. The multi-generation design process, where updated digital models informed subsequent

CCD optimization, showcases how DTs can bridge virtual simulations and real-world systems to enhance adaptability, efficiency, and robustness across a vehicle’s lifecycle. These findings demonstrate how the DT framework enables active suspension systems that co-evolve physical system design

with control strategies and self-improve through continual, data-driven learning.

Despite these promising results, the present study primarily focused on simulated environments with quantile learning of hypothetically real data. As a critical next step, future work will integrate the proposed framework with a fully instrumented physical suspension system, enabling experimental validation and data-driven model updating based on real-world measurements. Future work will also extend the framework to incorporate other advanced uncertainty quantification (UQ) and the formulation of robust or Pareto-optimal co-design problems that optimize a single suspension hardware configuration over mixed driving-style distributions. In addition, the current implementation targeted suspension co-design only, while future research will generalize the CCD formulation to the entire autonomous vehicle by co-optimizing interconnected physical subsystems (e.g., transmission gear ratios, tire–road friction coefficients, and engine performance) with driving control strategies. Such an integrated, system-level co-design will advance the development of adaptive, high-performance, and safe autonomous vehicles that leverage DTs as learning-based decision-making systems throughout their operational lifecycles.

Author contributions The authors' contribution is as follows: Y.-K. Tsai was responsible for the methodology, modeling, data analysis, software writing and original draft work. Y.-P. Chen and V. Karkaria contributed to result refinement and manuscript revisions. W. Chen supervised the research, provided guidance, and reviewed the manuscript.

Funding We are grateful for the grant support from the National Science Foundation (NSF)'s Engineering Research Center for Hybrid Autonomous Manufacturing: Moving from Evolution to Revolution (ERC-HAMMER) under Award Number EEC-2133630, NSF's FMRG: Manufacturing Advanced Electronics through Printing Using Bio-based and Locally Identifiable Compounds (MADE-PUBLIC), under Award Number CMMI-2037026, and NASA's Center for In-space Manufacturing (CISM-R2): Recycling and Regolith Processing under Award Number 80NSSC24M0176. Yi-Ping Chen also acknowledges the Taiwan-Northwestern Doctoral Scholarship to support his doctoral study.

Data availability All codes used in this paper are freely available through the [GitHub repository](#). The obtained results can be replicated by implementing the formulations and simulated solutions presented in this work. The data for producing the presented results can be made available by request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allison JT, Herber DR (2014) Special section on multidisciplinary design optimization: multidisciplinary design optimization of dynamic engineering systems. *AIAA J* 52(4):691–710
- Allison JT, Guo T, Han Z (2014) Co-design of an active suspension using simultaneous dynamic optimization. *J Mech Des* 136(8):081003
- Anh NT (2020) Control an active suspension system by using pid and lqr controller. *Int J Mech Prod Eng Res Develop* 10(3):7003–7012
- Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA (2017) Deep reinforcement learning: a brief survey. *IEEE Signal Process Mag* 34(6):26–38
- Asif FM, Roci M, Lieder M, Rashid A, Mihelič A, Kotnik S (2021) A methodological approach to design products for multiple lifecycles in the context of circular manufacturing systems. *J Clean Prod* 296:126534
- Asmat H, Din IU, Almogren A, Khan MY (2025) Digital twin with soft actor-critic reinforcement learning for transitioning from industry 4.0 to 5.0. *IEEE Access* 2025:1–10
- Bayat S, Allison JT (2025) Control co-design with varying available information applied to vehicle suspensions. *J Dyn Syst Meas Control*. <https://doi.org/10.1115/1.4069918>
- Buşoniü L, De Bruin T, Tolić D, Kober J, Palunko I (2018) Reinforcement learning for control: Performance, stability, and deep approximators. *Annu Rev Control* 46:8–28
- Chen Y-P, Chan K-Y (2021) Unknown parameter excitation and estimation for complex systems with dynamic performances. *J Mech Des* 143(9):091704
- Chen T, He Z, Ciocarlie M (2020) Hardware as policy: mechanical and computational co-optimization using deep reinforcement learning. Preprint at <http://arxiv.org/abs/2008.04460>
- Chen X, Huang D, Li M, Cai Y, Wen Z, Cai Z, Yang W (2023a) Evolving physical instinct for morphology and control co-adaptation. In: 2023 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 6616–6623
- Chen C, Xiang P, Lu H, Wang Y, Xiong R (2023b) C 2: Co-design of robots via concurrent-network coupling online and offline reinforcement learning. In 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 7487–7494. IEEE
- Chen Y-P, Wang L, Comlek Y, Chen W (2024) A latent variable approach for non-hierarchical multi-fidelity adaptive sampling. *Comput Methods Appl Mech Eng* 421:116773
- Chen Y-P, Karkaria V, Tsai Y-K, Rolark F, Quispe D, Gao RX, Cao J, Chen W (2025a) Real-time decision-making for digital twin in additive manufacturing with model predictive control using time-series deep neural networks. *J Manuf Syst* 80:412–424
- Chen Y-P, Tsai Y-K, Karkaria V, Chen W (2025b) Uncertainty-aware digital twins: Robust model predictive control using time-series deep quantile learning. *J Mech Des* 148(2):021702. <https://doi.org/10.1115/1.4069104> (<https://asmedigitalcollection.asme.org/mechanicaldesign/article-pdf/148/2/021702/7525270/md-25-1042.pdf>)

- Cui T, Allison JT, Wang P (2021) Reliability-based control co-design of horizontal axis wind turbines. *Struct Multidisc Optim* 64(6):3653–3679
- Deshmukh AP, Herber DR, Allison JT (2015) Bridging the gap between open-loop and closed-loop control in co-design: A framework for complete optimal plant and control architecture design. In: 2015 American control conference (ACC). IEEE, pp 4916–4922
- Dridi I, Hamza A, Ben Yahia N (2023) A new approach to controlling an active suspension system based on reinforcement learning. *Adv Mech Eng* 15(6):16878132231180480
- Dridi I, Hamza A, Yahia NB (2025) Optimizing full vehicle active suspension model with advanced reinforcement learning controller. *Int J Interact Des Manuf* 19(7):5269–5289
- Eimer T, Lindauer M, Raileanu R (2023) Hyperparameters in reinforcement learning and how to tune them. In: International conference on machine learning. PMLR, pp 9104–9149
- Fares A, Bani Younes A (2020) Online reinforcement learning-based control of an active suspension system using the actor critic approach. *Appl Sci* 10(22):8060
- Fathy HK, Reyer JA, Papalambros PY, Ulsov A (2001) On the coupling between the plant and controller optimization problems. In: Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148), vol. 3. IEEE, pp 1864–1869
- Garcia-Sanz M (2019) Control co-design: an engineering game changer. *Adv Control Appl Eng Ind Syst* 1(1):e18
- Go TF, Wahab DA, Hishamuddin H (2015) Multiple generation life-cycles for product sustainability: the way forward. *J Clean Prod* 95:16–29
- Goodarzi A, Khajepour A (2017) Vehicle suspension system technology and design. Springer
- Grievies M (2014) Digital twin: manufacturing excellence through virtual factory replication. *White Paper* 1(2014):1–7
- Grievies M, Vickers J (2017) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. Springer International Publishing, Cham, pp 85–113
- Gupta A, Savarese S, Ganguli S, Fei-Fei L (2021) Embodied intelligence via learning and evolution. *Nat Commun* 12(1):5721
- He T, Su W (2023) Co-design of active vibration control and optimal sensor and actuator placement for a flexible wing using reinforcement learning. *Proc Inst Mech Eng G J Aerosp Eng* 237(10):2240–2251
- Herber DR, Allison JT (2019) Nested and simultaneous solution strategies for general combined plant and control design problems. *J Mech Des* 141(1):011402
- Hu C, Hu Z, Zheng P, Kim T, González VA, San O (2023) Special issue on advanced optimization enabling digital twin technology. *Struct Multidisc Optim* 66(10):218
- Hu Z, Hu C, Hu W (2024) A tutorial on digital twins for predictive maintenance. In: Structural health monitoring/management (SHM) in aerospace structures, pp 453–501
- Jin IM, Schmit LA (1992) Control design variable linking for optimization of structural/control systems. *AIAA J* 30(7):1892–1900
- Karkaria V, Goeckner A, Zha R, Chen J, Zhang J, Zhu Q, Cao J, Gao RX, Chen W (2024) Towards a digital twin framework in additive manufacturing: Machine learning and bayesian optimization for time series process optimization. *J Manuf Syst* 75:322–332
- Karkaria V, Chen J, Luey C, Siuta C, Lim D, Radulescu R, Chen W (2025a) A digital twin framework utilizing machine learning for robust predictive maintenance: enhancing tire health monitoring. *J Comput Inf Sci Eng* 25(7):071003
- Karkaria V, Tsai Y-K, Chen Y-P, Chen W (2025b) An optimization-centric review for integrating artificial intelligence and digital twin technologies in manufacturing. *Eng Optim*. <https://doi.org/10.1080/0305215X.2024.2434201>
- Karkaria V, Chen Y-P, Tsai Y-K, Erdogan G, Radulescu R, Chen W (2025c) Ai-driven digital twin framework for predictive maintenance: Dynamic decision making and adaptive model updates. In: International design engineering technical conferences and computers and information in engineering conference. American Society of Mechanical Engineers
- Ladosz P, Weng L, Kim M, Oh H (2022) Exploration in deep reinforcement learning: a survey. *Inf Fusion* 85:1–22
- Lee GW, Hyun M, Kang DO, Heo SJ (2022) High-efficiency active suspension based on continuous damping control. *Int J Automot Technol* 23(1):31–40
- Lee D, Jin S, Lee C (2022b) Deep reinforcement learning of semi-active suspension controller for vehicle ride comfort. *IEEE Trans Veh Technol* 72(1):327–339
- Li L, Li H, Wang R, Liu Y, Wang G, Yan Y (2025) Digital twin structural health monitoring driven by multi-fidelity time-series surrogate models. *J Ind Inf Integr* 2025:100918
- Luck KS, Amor HB, Calandra R (2020) Data-efficient co-adaptation of morphology and behaviour with deep reinforcement learning. In: Conference on Robot Learning. PMLR, pp. 854–869
- Ma P, Du T, Zhang JZ, Wu K, Spielberg A, Katzschmann RK, Matusik W (2021) Diffaqua: a differentiable computational design pipeline for soft underwater swimmers with shape interpolation. *ACM Trans Graph* 40(4):1–14
- Manna S, Mani G, Ghildiyal S, Stonier AA, Peter G, Ganji V, Murugesan S (2022) Ant colony optimization tuned closed-loop optimal control intended for vehicle active suspension system. *IEEE Access* 10:53735–53745
- Marykovskiy Y, Clark T, Deparday J, Chatzi E, Barber S (2024) Architecting a digital twin for wind turbine rotor blade aerodynamic monitoring. *Front Energy Res* 12:1428387
- Ming L, Yibin L, Xuewen R, Shuaishuai Z, Yanfang Y (2020) Semi-active suspension control based on deep reinforcement learning. *IEEE Access* 8:9978–9986
- Nag U, Sharma SK, Kumar V (2022) Multiple life-cycle products: a review of antecedents, outcomes, challenges, and benefits in a circular economy. *J Eng Des* 33(3):173–206
- Nash AL, Pangborn HC, Jain N (2021) Robust control co-design with receding-horizon mpc. In: 2021 American control conference (ACC). IEEE, pp 373–379
- National Academies of Sciences, Engineering, and Medicine (2023) Foundational research gaps and future directions for digital twins. National Academies Press, Washington
- Nguyen DN, Nguyen TA (2023) The dynamic model and control algorithm for the active suspension system. *Math Probl Eng* 2023(1):2889435
- Onaji I, Tiwari D, Soulatiantork P, Song B, Tiwari A (2022) Digital twin in manufacturing: conceptual framework and case studies. *Int J Comput Integr Manuf* 35(8):831–858
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L (2019) Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 32:2
- Qiu H, Al-Nussairi AKJ, Chevinli ZS, Singh N, Singh S, Chyad MH, Yu J, Maesoumi M (2025) Integrating digital twins with neural networks for adaptive control of automotive suspension systems. *Sci Rep* 15(1):11078
- Rosa LO, Branco CTNM (2024) Digital twin-enabled fault detection for suspension systems in autonomous mining haulage vehicles. Technical report, SAE Technical Paper
- Sato Y, Yuhn C, Kobayashi H, Kawamoto A, Nomura T (2025) Computational co-design of structure and feedback controller for locomoting soft robots. *Struct Multidisc Optim* 68(9):1–21
- Schaff C, Yunis D, Chakrabarti A, Walter MR (2019) Jointly learning to construct and control agents using deep reinforcement learning. In: 2019 international conference on robotics and automation (ICRA). IEEE, pp 9798–9805

- Schmid M, Rath D, Diebold U (2022) Why and how Savitzky-Golay filters should be replaced. *ACS Meas Sci Au* 2(2):185–196
- Schulman J, Levine S, Abbeel P, Jordan M, Moritz P (2015) Trust region policy optimization. In: International conference on machine learning. PMLR, pp 1889–1897
- Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. Preprint at <http://arxiv.org/abs/1707.06347>
- Semeraro C, Lezoche M, Panetto H, Dassisti M (2021) Digital twin paradigm: a systematic literature review. *Comput Ind* 130:103469
- Shakya AK, Pillai G, Chakrabarty S (2023) Reinforcement learning algorithms: a brief survey. *Expert Syst Appl* 231:120495
- Sisson W, Karve P, Mahadevan S (2022) Digital twin for component health-and stress-aware rotorcraft flight control. *Struct Multidisc Optim* 65(11):318
- Soliman A, Kaldas M (2021) Semi-active suspension systems from research to mass-market-a review. *J Low Freq Noise Vib Act Control* 40(2):1005–1023
- Sun W, Gao H, Shi P (2020) Advanced control for vehicle active suspension systems, vol 204. Springer
- Sun J, Yao M, Xiao X, Xie Z, Zheng B (2023) Co-optimization of morphology and behavior of modular robots via hierarchical deep reinforcement learning. In: Robotics: science and systems
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction, vol 1. MIT Press, Cambridge
- Thelen A, Zhang X, Fink O, Lu Y, Ghosh S, Youn BD, Todd MD, Mahadevan S, Hu C, Hu Z (2022) A comprehensive review of digital twin—part 1: modeling and twinning enabling technologies. *Struct Multidisc Optim* 65(12):354
- Thelen A, Zhang X, Fink O, Lu Y, Ghosh S, Youn BD, Todd MD, Mahadevan S, Hu C, Hu Z (2023) A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives. *Struct Multidisc Optim* 66(1):1
- Tsai Y-K, Malak RJ (2022a) A constraint-handling technique for parametric optimization and control co-design. In: International design engineering technical conferences and computers and information in engineering conference. American Society of Mechanical Engineers
- Tsai Y-K, Malak RJ (2022b) Design of approximate explicit model predictive controller using parametric optimization. *J Mech Des* 144(12):1
- Tsai Y-K, Malak RJ (2023) Robust control co-design using tube-based model predictive control. In: 2023 American control conference (ACC). IEEE, pp 769–775
- Tsai Y-K, Malak RJ (2024) Surrogate-assisted constraint-handling technique for parametric multi-objective optimization. *Struct Multidisc Optim* 67(9):167
- Tsai Y-K, Malak RJ (2025) Control co-design with performance-robustness trade-off using tube-based stochastic model predictive control. *ASME Letters in Dynamic Systems and Control*
- Tsai Y-K, Malak RJ (2026) Parametric multi-objective optimization for simultaneous and nested control co-design formulations with tube-based model predictive controllers. *J Mech Des*
- Tsai Y-K, Karkaria V, Chen Y-P, Chen W (2025a) Digital twin-enabled multi-generation control co-design with deep reinforcement learning. arXiv preprint [arXiv:2510.10694](https://arxiv.org/abs/2510.10694)
- Tsai Y-K, Karkaria V, Chen Y-P, Chen W (2025b) Multi-generation control co-design for digital twin-enabled systems with deep reinforcement learning. In: International design engineering technical conferences and computers and information in engineering conference. American Society of Mechanical Engineers
- van Beek A, Karkaria V, Chen W (2023) Digital twins for the designs of systems: a perspective. *Struct Multidisc Optim* 66(3):49
- van Diepen M, Shea K (2022) Co-design of the morphology and actuation of soft robots for locomotion. *J Mech Des* 144(8):083305
- Wang T, Zhou Y, Fidler S, Ba J (2019) Neural graph evolution: automatic robot design. *Int Conf Learn Represent*. <https://doi.org/10.48550/arXiv.1906.05370>
- Wang Y, Liu A, Tao F, Nee A (2020) Digital twin driven conceptual design. *Digital twin driven smart design*. Elsevier, Amsterdam, pp 33–66
- Wang X, Wang S, Liang X, Zhao D, Huang J, Xu X, Dai B, Miao Q (2022) Deep reinforcement learning: a survey. *IEEE Trans Neural Netw Learn Syst* 35(4):5064–5078
- Wang Y, Wu S, Zhang T, Chang Y, Fu H, Fu Q, Wang X (2023) Preco: Enhancing generalization in co-design of modular soft robots via brain-body pre-training. In: Conference on robot learning. PMLR, pp 478–498
- Wang C, Cui X, Zhao S, Zhou X, Song Y, Wang Y, Guo K (2024) A deep reinforcement learning-based active suspension control algorithm considering deterministic experience tracing for autonomous vehicle. *Appl Soft Comput* 153:111259
- Weaver-Rosen JM, Leal PB, Hartl DJ, Malak RJ Jr (2020) Parametric optimization for morphing structures design: application to morphing wings adapting to changing flight conditions. *Struct Multidisc Optim* 62(6):2995–3007
- Wexler B, Sarafian E, Kraus S (2022) Analyzing and overcoming degradation in warm-start reinforcement learning. In 2022 IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp 4048–4055
- Xu H, Carrillo L (2017) Fast reinforcement learning based distributed optimal flocking control and network co-design for uncertain networked multi-uav system. In: Unmanned systems technology XIX, vol 10195. pp 347–355, SPIE
- Yu G, Wang Y, Mao Z, Hu M, Sugumaran V, Wang YK (2021) A digital twin-based decision analysis framework for operation and maintenance of tunnels. *Tunn Undergr Space Technol* 116:104125
- Yuhn C, Sato Y, Kobayashi H, Kawamoto A, Nomura T (2023) 4d topology optimization: integrated optimization of the structure and self-actuation of soft bodies for dynamic motions. *Comput Methods Appl Mech Eng* 414:116187

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.