Expert Systems With Applications 287 (2025) 128166

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

A survey on physics informed reinforcement learning: Review and open problems

Chayan Banerjee [a,*], Kien Nguyen [a], Clinton Fookes [a], Maziar Raissi [b]

[a] Queensland University of Technology, 2 George Str., 4000 QLD, Brisbane, Australia
[b] University of Colorado, 2000 Colorado Ave., Boulder, 80309 CO, USA

ARTICLE INFO

Keywords:
Physics-informed, Reinforcement learning, Machine learning, Neural network, Deep learning, Information fusion

ABSTRACT

The fusion of physical information in machine learning frameworks has revolutionized many application areas. This involves enhancing the learning process by incorporating physical constraints and adhering to physical laws. This work explores their utility for reinforcement learning applications. A thorough review of the literature on the fusion of physics information or physics priors in reinforcement learning approaches, commonly referred to as physics-informed reinforcement learning (PIRL), is presented. A novel taxonomy is introduced with the reinforcement learning pipeline as the backbone to classify existing works, compare and contrast them, and derive crucial insights. Existing works are analyzed with regard to the representation/form of the governing physics modeled for integration, their specific contribution to the typical reinforcement learning architecture, and their connection to the underlying reinforcement learning pipeline stages. Core learning architectures and physics incorporation biases (i.e., observational, inductive, and learning) of existing PIRL approaches are identified and used to further categorize the works for better understanding and adaptation. By providing a comprehensive perspective on the implementation of the physics-informed capability, the taxonomy presents a cohesive approach to PIRL. It identifies the areas where this approach has been applied, as well as the gaps and opportunities that exist. Additionally, the review highlights unresolved issues and challenges, while also incorporating potential and emerging solutions to guide future research. This nascent field holds great potential for enhancing reinforcement learning algorithms by increasing their physical plausibility, precision, data efficiency, and applicability in real-world scenarios.

1. Introduction

Through trial-and-error interactions with the environment, Reinforcement Learning (RL) offers a promising approach to solving decision-making and optimization problems. Over the past few years, RL has accomplished impressive feats in handling difficult tasks, in such domains as autonomous driving (Chen, Yuan, & Tomizuka, 2019; Toromanoff, Wirbel, & Moutarde, 2020), locomotion control (Peng, Berseth, Yin, & Van De Panne, 2017; Xie, Berseth, Clary, Hurst, & van de Panne, 2018), robotics (Levine, Finn, Darrell, & Abbeel, 2016; Neunert et al., 2020), continuous control (Banerjee, Chen, & Noman, 2022a,b; Banerjee, Chen, Noman, & Zamani, 2022c), and multi-agent systems and control (Chen, Lewis, Xie, Xie, & Liu, 2020; Gao & Pavel, 2021). A majority of these successful approaches are purely data-driven and leverage trial-and-error to freely explore the search space. RL methods work well in simulations, but they struggle with real-world data because of the disconnection between simulated setups and the complexities of real world systems. Major RL challenges (Dulac-Arnold et al., 2021), that are consistently addressed in latest research includes sample efficiency (Barth-Maron et al., 2018; Mnih et al., 2013), high dimensional continuous state and action spaces (Dulac-Arnold et al., 2015; Tessler, Zahavy, Cohen, Mankowitz, & Mannor, 2019), safe exploration (Garcia & Fernández, 2012; Gu et al., 2022), multi-objective and well-defined reward function (Booth et al., 2023; Knox, Allievi, Banzhaf, Schmitt, & Stone, 2023), perfect simulators and learned model (Cutler, Walsh, & How, 2015; Osiński et al., 2020) and policy transfer from offline pre-training (Levine, Kumar, Tucker, & Fu, 2020; Yang & Nachum, 2021).

When it comes to machine learning, incorporating mathematical physics into the models can lead to more meaningful solutions. This approach, known as physics-informed machine learning, helps neural networks learn from incomplete physics information and imperfect data more efficiently, resulting in faster trainingtimes and better

* Corresponding author.
E-mail addresses: c.banerjee@qut.edu.au (C. Banerjee), k.nguyenthanh@qut.edu.au (K. Nguyen), c.fookes@qut.edu.au (C. Fookes), maziar.raissi@colorado.edu (M. Raissi).

https://doi.org/10.1016/j.eswa.2025.128166
Received 21 December 2023; Received in revised form 2 April 2025; Accepted 7 May 2025
Available online 14 May 2025

generalization. Additionally, it can assist in tackling high dimensionality applications and ensure that the resulting solution is physically sound and follows the underlying physical law (Banerjee, Nguyen, Fookes, & Karniadakis, 2023; Hao et al., 2022; Karniadakis et al., 2021). Among the various sub-fields of ML, RL is the natural candidate for incorporating physics information since most RL-based solutions deal with real-world problems and have an explainable physical structure.

Recent research has seen substantial improvement in addressing the RL challenges by incorporating physics information in the training pipeline. For example, PIRL approaches seek to use physics to reduce high-dimensional continuous states with intuitive representations and better simulation. A low-dimensional representation adhering to physical model PDEs is learned in Gokhale, Claessens, and Develder (2022), while (Cao et al., 2023a) uses features from a supervised surrogate model. Learning a good world model is a quicker and safer alternative to training RL agents in the real world. Ramesh and Ravindran (2023) incorporate physics into the network for better world models, and Xie, Patil, Moldovan, Levine, and Abbeel (2016) utilize a high-level specification robot morphology and physics for rapid model identification.

A well-defined reward function is crucial for successful reinforcement learning, PIRL approaches also seek to incorporate physical constraints into the design for safe learning and more efficient reward functions. For example, in (Korivand, Jalili, & Gong, 2023) the designed reward incorporates IMU sensor data, imbibing inertial constraints, while in (Li, He, Li, Shi, & Zeng, 2023) the physics informed reward is designed to satisfy explicit operational targets. To ensure safe exploration during training and deployment, works such as Yang, Jiang, Liu, Chen, and Li (2023); Zhao, Wang, and Yue (2023) learn a data-driven barrier certificate based on physical property-based losses and a set of unsafe state vectors.

There are several lines of PIRL research dedicated to exploring more efficient exploration of the search space and effective policy deployment for real-world systems. Some approaches were developed to improve simulators for sample efficiency and better sim to real transfer (Alam, Shtein, Barton, & Hoelzle, 2021; Lowrey, Kolev, Dao, Rajeswaran, & Todorov, 2018). Carefully selecting task-specific state representations (Han et al., 2022; Jurj et al., 2021), reward functions (Cao, Mao, Sha, & Caccamo, 2023b,c), and action spaces (Wang, 2022; Zhao et al., 2023) has been shown to improve both the time to convergence and performance. To sum it up, integrating underlying physics about the learning task structure has been found to improve performance and accelerate convergence.

This survey paper draws on high-quality sources, including Semantic Scholar, Google Scholar, IEEE Xplore, and SpringerLink. Keywords like *physics-informed, physics-aided, physics informed reinforcement learning,* and *physics priors*, were used to search for relevant peer-reviewed journals, conference papers, and technical reports, ensuring a comprehensive review. Fig. 1 illustrates the volume of research works identified in this process.

PIRL has been a growing trend in the literature, as demonstrated in the increasing number of papers published in this area over the past seven years, as shown in Fig. 1. The bar chart indicates that this field is gaining more attention, and we can anticipate even more in the future. Our contributions in this paper are summarized as follows:

1) *Taxonomy:* We propose a unified taxonomy to investigate what physics knowledge/processes are modeled, how they are represented, and the strategies to incorporate them into RL approaches.
2) *Algorithmic Review:* We present state-of-the-art approaches on the physics information guided/ physics informed RL methods, using unified notations, simplified functional diagrams and discussion on latest literature.
3) *Training and evaluation benchmark Review:* We analyze the evaluation benchmarks used in the reviewed literature, thus presenting popular evaluation and benchmark platforms/ suites for understanding the popular trend and also for easy reference.

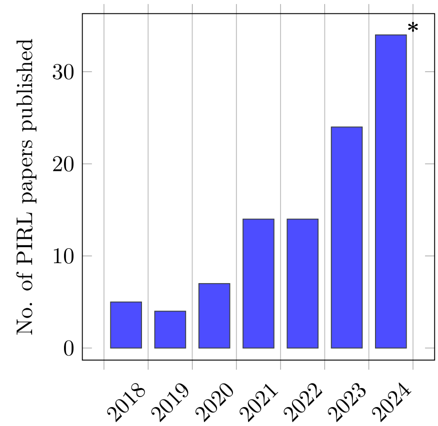


Fig. 1. PIRL papers published over the years. This graph illustrates the exponential growth of PIRL papers over the last seven years. *The* 2024* *data point indicates the trend, with only key 2024 contributions highlighted in the Emerging Trends section in Sec:4.3.3.*

4) *Analysis:* We delve deep into a wide range of model based and model free RL applications over diverse domains. We analyze in detail how physics information is integrated into specific RL approaches, what physical processes have been modeled and incorporated, and what network architectures or network augmentations have been utilized to incorporate physics.
5) *Open Problems:* We summarize our perspectives on the challenges, open research questions and directions for future research.

Difference from Other survey papers: Karniadakis et al. (2021) provided one of the most comprehensive reviews on machine learning (ML) in the context of physics-informed (PI) methods, but approaches in the RL domain has not been discussed. The work by Hao et al. (2022) also provided an overview of physics-informed machine learning, where the authors briefly touch upon the topic of PIRL. Another recent study by EEßerer, Bach, Jestel, Urbann, and Kerner (2022) showcased the use of prior knowledge to guide reinforcement learning (RL) algorithms, specific to robotic applications. The authors categorize knowledge into three types: expert knowledge, world knowledge, and scientific knowledge. Our paper offers a focused and comprehensive review specially on the RL approaches that utilize the structure, properties, or constraints unique to the underlying physics of a process/system. Our scope of application domains is not limited to robotics, but also spanning to motion control, molecular structure optimization, safe exploration, and robot manipulation.

The rest of this paper is organized as follows. In Section 2, we provide a brief overview of the Physics informed ML paradigm. In Section 3, we present RL fundamentals/ framework in Section 3.1 and provide a definition with an intuitive introduction to PIRL in section 3.2. Most importantly we introduce a comprehensive taxonomy in Section 3.3 threading together physics information types, PIRL methods that implement those information and RL pipeline as a backbone. Later in Section 3.4 we present and elaborate on two additional categories: Learning architecture and Bias, through which the implementation side of the literature is explained more precisely. In Section 4 we present an elaborate review and analysis of latest PIRL literature. In Section 5, we discuss the different open problems, challenges and research directions that may be addresses in future works by interested researchers. Finally Section 6 concludes the paper. We have also included a list of abbreviations used; see Table 1.

2. Physics-informed machine learning (PIML): An overview

The aim of PIML is to merge mathematical physics models and observational data seamlessly in the learning process. This helps to guide the

Table 1
A list of abbreviations used in this article.

FSA	Finite State Automata
FEA	Finite Element Analysis
CFD	Computational Fluid Dynamics
MDP	Markov Decision Process
MBRL	Model based Reinforcement Learning
MFRL	Model Free Reinforcement Learning
CBF	Control Barrier Function
CBC	Control Barrier Certificate
NBC	Neural Barrier Certificate
CLBF	Control Lyapunov Barrier Function
NBC	Neural Barrier Certificate
DFT	Density Functional Theory
AC	Actor Critic
MPC	Model Predictive Control
DDP	Differential Dynamic Programming
NPG	Natural Policy Gradient
TL	Temporal Logic
DMP	Dynamic Movement Primitive
WBTG	Whole Body Trajectory Generator
DPG	Deterministic Policy Gradient
DPPO	Distributed proximal Policy optimization
ABM (PPR)	Adjoint based method
APG	Analytic Policy Gradient
WBIC	Whole Body Impulse Controller
LNN	Lagrangian Neural Network

process towards finding a physically consistent solution even in complex scenarios that are partially observed, uncertain, and high-dimensional (Cuomo et al., 2022; Hao et al., 2022; Kashinath et al., 2021). Adding physics knowledge to machine learning models has numerous benefits, as discussed in Kashinath et al. (2021); Meng, Seo, Cao, Griesemer, and Liu (2022a). For example, Zhang et al. (2024) introduced a traffic state estimation model which combines computational graph with physics-informed deep learning methods, based on traffic flow model. Again in Zheng et al. (2023) the authors integrated prior structures from traffic science, into a general sequential prediction model, for improved network congestion prediction. As seen in examples above, the physics information captures the vital physical principles of the process being modeled and brings following advantages

1. Ensures that the ML model is consistent both physically and scientifically.
2. Increases data efficiency in model training, meaning that the model can be trained with fewer data inputs.
3. Accelerates the model training process, allowing models to converge faster to an optimal solution.
4. Increases the generalizability of trained models, enabling them to make better predictions for scenarios that were not seen during the training phase.
5. Enhances the transparency and interpretability of models, making them more trustworthy and explainable.

According to literature, there are three strategies for integrating physics knowledge or priors into machine learning models: observational bias, learning bias, and inductive bias.

Observational bias: This approach uses multi-modal data that reflects the physical principles governing their generation (Kashefi, Rempe, & Guibas, 2021; Li et al., 2020; Lu, Jin, Pang, Zhang, & Karniadakis, 2021; Yang & Perdikaris, 2019). The deep neural network (DNN) is trained directly on observed data, with the goal of capturing the underlying physical process. The training data can come from various sources such as direct observations, simulation or physical equation-generated data, maps, or extracted physics data induction.

Learning bias: One way to reinforce prior knowledge of physics is through soft penalty constraints. This approach involves adding extra terms to the loss function that are based on the physics of the process, such as momentum or conservation of mass. An example of this is physics-informed neural networks (PINN), which combine information from measurements and partial differential equations (PDEs) by embedding the PDEs into the neural network's loss function using automatic differentiation (Karniadakis et al., 2021). Some prominent examples of soft penalty based approaches includes statistically constrained GAN (Wu et al., 2020), physics-informed auto-encoders (Erichson, Muehlebach, & Mahoney, 2019) and encoding invariances by soft constraints in the loss function InvNet (Shah et al., 2019).

Inductive bias: Custom neural network-induced "hard" constraints can incorporate prior knowledge into models. For instance, Hamiltonian NN (Greydanus, Dzamba, & Yosinski, 2019) draws inspiration from Hamiltonian mechanics and trains models to respect exact conservation laws, resulting in better inductive biases. Lagrangian Neural Networks (LNNs) (Cranmer et al., 2020) introduced by Cranmer et al. can parameterize arbitrary Lagrangians using neural networks, even when canonical momenta are unknown or difficult to compute. Meng, Yang, Mao, del Águila Ferrandis, and Karniadakis (2022b) uses a Bayesian framework to learn functional priors from data and physics with a PI-GAN, followed by estimating the posterior PI-GAN's latent space using the Hamiltonian Monte Carlo (HMC) method. Additionally, DeepONets (Lu et al., 2021) networks are used in PDE agnostic physical problems.

3. PIRL: Fundamentals, taxonomy and examples

In this section, we will explain how physics information can be integrated into reinforcement learning applications.

3.1. RL fundamentals

RL algorithms use reward signals from the environment to learn the best strategy to solve a task through trial and error. They effectively solve sequential decision-making problems that follow the Markov Decision Process (MDP) framework. In the RL paradigm, there are two main players: the agent and the environment. The environment refers to the world where the agent resides and interacts. Through agent-environment interactions the agent perceives the state of the world and decides on the appropriate action to take.

The agent-environment RL framework, see Fig. 2, is a large abstraction of the problem of goal-directed learning from interaction (Sutton & Barto, 2018). The details of control apparatus, sensors, and memory are abstracted into three signals between the agent and the environment: the control/ action, the state and the reward. Though typically, the agents computes the rewards, but by the current convention anything that cannot be changed arbitrarily by the agent is considered outside of it and hence the reward function is shown as a part of the environment.

MDP is typically represented by the tuple $(\mathcal{S}, \mathcal{A}, R, P, \gamma)$, where $\mathcal{S}$ represents the states of the environment, $\mathcal{A}$ represents set of actions that the RL agent can take. Reward function may be typically represented as $R(s_{t+1}, a_t)$ a function of next state and current action. The function generates the reward due to action induced state transition from s_t to s_{t+1}. $P(s_{t+1}|s_t, a_t)$ is the environment model that returns the probability of transitioning to state s_{t+1} from s_t. Finally the discount factor $\gamma \in [0, 1]$, determines the amount of emphasis given to the immediate rewards relative to that of future rewards.

The RL framework typically organizes the agent's interactions with the environment into episodes. In each episode, the agent starts at a particular initial state s_1 sampled from an initial distribution $p(s_1)$, which

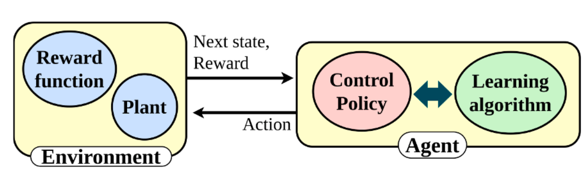


Fig. 2. Agent-environment framework, of RL paradigm. Here the reward generating function and the system/ plant is abstracted as the environment. And the control policy (e.g. a DNN) and the learning algorithm, forms the RL agent.

is part of the state space $\mathcal{S}$ of the MDP. At each timestep t, the agent observes the current state $s_t \in \mathcal{S}$ and samples an action $a_t \in \mathcal{A}$ from its latest policy $\pi_\phi(a_t|s_t)$ based on the state s_t, where ϕ represents the policy parameters. The action space of the MDP is denoted by $\mathcal{A}$.

Next, the agent applies the action a_t into the environment, which results in a new state s_{t+1} given by the dynamics of the MDP, i.e., $s_{t+1} \sim p(s_{t+1}|s_t, a_t)$. The agent also receives a reward $r_t = R(s_{t+1}, a_t)$, which can be construed as the desirability of a certain state transition from the context of the given task.

The above process is repeated up to a certain time horizon T, which may also be infinite. The agent-environment interaction is recorded as a trajectory, and the closed-loop trajectory distribution for the episode $t = 1, \cdots, T$ can be represented by,

$$p_\phi(\tau) = p_\phi(s_1, a_1, s_2, a_2, \cdots, s_T, a_T, s_{T+1}) \tag{1}$$

$$= p(s_1) \prod_{t=1}^{T} \pi_\phi(a_t|s_t) p(s_{t+1}|s_t, a_t), \tag{2}$$

where $\tau = (s_1, a_1, s_2, a_2, \cdots, s_T, a_T, s_{T+1})$ represents the sequence of states and control actions. The objective is to find an optimal policy represented by the parameter,

$$\phi^* = \arg\max_\phi \underbrace{\mathbf{E}_{\tau\sim p_\phi(\tau)} \left[\sum_{t=1}^{T} \gamma^t R(a_t, s_{t+1}) \right]}_{\mathcal{J}(\phi)}, \tag{3}$$

which maximizes the objective function $\mathcal{J}(\phi)$, γ is a parameter called discount factor, where $0 \le \gamma \le 1$. γ determines the present value of the future reward, i.e., a reward received at k timesteps in the future is worth only γ^{k-1} times what it would be worth if received immediately.

Model-free and model-based RL: In RL, algorithms can be classified based on whether the environment model is available during policy optimization. The environment dynamics are represented as $p(s_{t+1}, r_t) = Pr(s_{t+1}, r_t|s_t, a_t)$, which means that given a state and action, the environment model can predict the state transition and the corresponding reward. Access to an environment model allows the agent to plan and choose between options and also improves sample efficiency compared to model-free approaches. However, the downside is that the environment's groundtruth model is typically not available, and learning a perfect model of the real world is challenging. Additionally, any bias in the learned model can lead to good performance in the learned model but poor performance in the real environment.

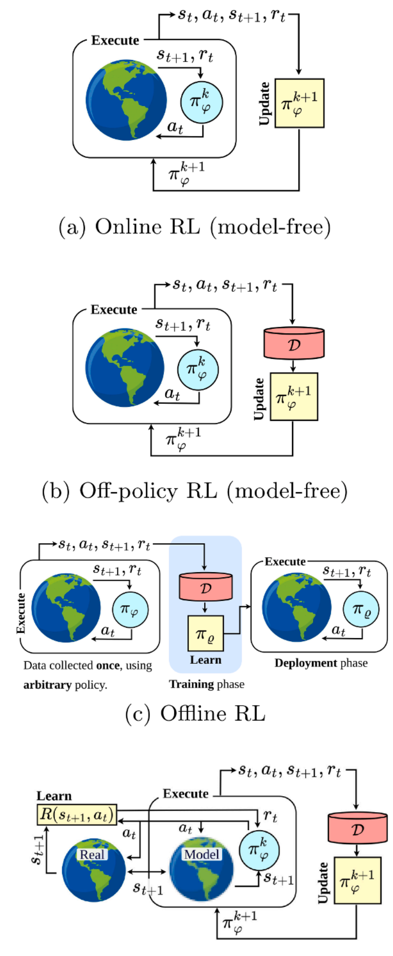
Online, Off-policy and Offline RL: Online RL algorithms, e.g. PPO, TRPO, and A3C, optimize policies by using only data collected while following the latest policy, creating an approximator for the state or action value functions, used to update the policy. Off-policy RL algorithms, e.g. SAC, TD3 and IPNS, involve the agent updating its policy and other networks using data collected at any point during training. This data is stored in a buffer called the experience replay buffer and is in the form of tuples. Mini-batches are sampled from the buffer and used for the training process. Offline RL algorithms use a fixed dataset called $\mathcal{D}$ collected by a policy π_ζ to learn the optimal policy. This allows for the use of large datasets collected previously.

Combining model-free/model-based with online/off-policy/offline categorization, typical RL architectures can be presented as Fig. 3.

3.2. Introduction

3.2.1. Definition of PIRL

The concept of physics-informed RL involves incorporating physics structures, priors, and real-world physical variables into the policy learning or optimization process. Physics induction helps improve the effectiveness, sample efficiency and accelerated training of RL algorithms/ approaches, for complex problem-solving and real-world deployment. Depending on the specific problem or scenario, different physics priors can be integrated using various RL methods at different stages of the RL framework Fig. 4.



(a) Online RL (model-free)

(b) Off-policy RL (model-free)

(c) Offline RL

(d) Off-policy RL (model-based)

Fig. 3. Typical RL architectures, based on model use and interaction with the environment.

3.2.2. Intuitive introduction to physics priors in RL

Physics priors come in different forms, like intuitive physical rules or constrains, underlying mathematical/ guiding equations and physics simulators, to name a few. Here we discuss a couple of intuitive examples. In Xie et al. (2016), the physical characteristics of the system were utilized as priors. The high-level specifications of a robot's morphology such as the number and connectivity structure of links were used as physics priors. This feature based representation of the system dynamics enabled rapid model identification in this model based RL setup. In another example, pertaining to adaptive cruise control problem, Jurj et al. (2021) (see Fig. 5), physics information in the form of "jam-avoiding distance" (based on desired physical parameters e.g. velocity and acceleration constraints, minimum jam avoiding distance etc.) is included in state space input to the RL agent. Physics info. incorporation results in a RL controller which performs with less collisions and enables more equidistant travel.

3.2.3. PIRL vs Conventional RL

PIRL fundamentally differs from conventional RL through explicit incorporation of physical principles. While traditional RL relies solely on data-driven environment interactions, PIRL integrates physical laws through various mechanisms including reward design (Korivand et al., 2023; Li et al., 2023; Siekmann, Godse, Fern, & Hurst, 2021), state representation (Jurj et al., 2021; Shi et al., 2023), action constraints (Cheng, Orosz, Murray, & Burdick, 2019; Li et al., 2021), and specialized model architectures (Bahl, Mukadam, Gupta, & Pathak, 2020; Ramesh & Ravindran, 2023). This physics-guided approach yields significant

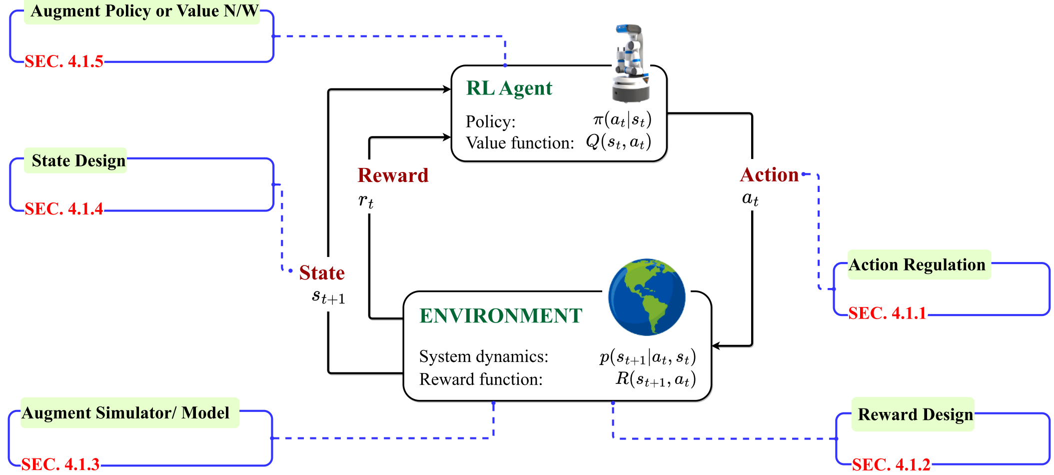


Fig. 4. Map of physics incorporation (PI) in the conventional Reinforcement Learning (RL) framework.

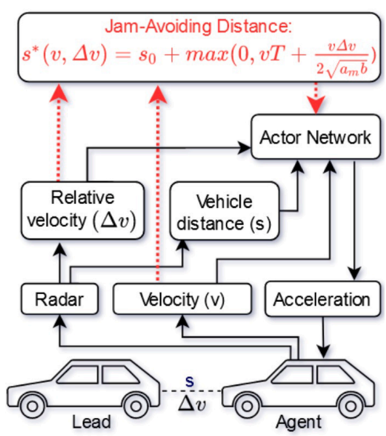


Fig. 5. An illustrative example of physics incorporation in RL application, Jurj et al. (2021). Here the RL agent is fed with an additional state variable: jam avoiding distance, which is based on desired physical parameters and primary state variables.

advantages: (1) improved sample efficiency by reducing exploration needs (Golemo, Taiga, Courville, & Oudeyer, 2018; Lowrey et al., 2018), (2) enhanced model accuracy through physically-constrained dynamics (Gao, Chen, Li, & Zhang, 2022), (3) better generalization to unseen scenarios (Rodwell & Tallapragada, 2023; Veerapaneni et al., 2020), (4) reduced model bias in model-based approaches (Lee, Seo, Lee, Lee, & Shin, 2020), and (5) physically plausible policies (Cao et al., 2023b; Wang, 2022). PIRL excels in complex dynamic systems (robotics, fluid dynamics) and safety-critical applications (Ohnishi, Wang, Notomista, & Egerstedt, 2019; Zhao et al., 2023) where unconstrained exploration could lead to dangerous states and where small errors can have significant consequences.

3.3. PIRL taxonomy

We introduce three categories based on the type of physics information, the methods used for its augmentation, and the stages of the reinforcement learning (RL) pipeline where it is integrated. These categories which forms the *PIRL Taxonomy,* are explained below with specific examples of their implementation. In Fig. 9 the implementation pathways are visualized, which shows how physics information flows through to PIRL methods and finally integrates into the RL pipeline.

3.3.1. Physics information (types): Representation of physics priors

There are different types/ forms of physics information, e.g. mathematical representation of the physical system like PDE/ODE and physics enriched simulators. Based on the type of the physics information representation, works can be typically categorized as follows.

1) *Differential and algebraic equations (DAE):* Many works use system dynamics representations, such as partial/ordinary differential equations (PDE/ ODE) and boundary conditions (BC), as physics priors primarily through PINN and other special networks. For example in transient voltage control application (Gao et al., 2022), a PINN is trained using PDE of transient process. The PINN learns a physical constraint which it transfers to the loss term of the RL algorithm.
2) *Barrier certificate and physical constraints (BPC):* It is imperative to regulate agent exploration in safety-critical applications of reinforcement learning. One way it is addressed in recent research is through the use of optimization-based control theoretic constraints. Use of concepts like control Lyapunov function (CLF) (Choi, Castaneda, Tomlin, & Sreenath, 2020; Li & Belta, 2019), barrier certificate/ barrier function (BF), control barrier function/ certificate (CBF/ CBC) (Cai, Cao, Lu, Zhang, & Xiong, 2021; Cheng et al., 2019) is made in recent safety critical RL applications. Barrier certificate is generally used to establish a safe set of desired states for a system. A control barrier function is then employed to devise a control law that keeps the states within the safety set. In certain scenarios barrier functions are represented as NNs and learned through data driven approaches (Zhao, Zhang, & Li, 2022; Zhao et al., 2023). In above control theoretic approaches the system dynamics either partial or learnable and safety sets represent the primary physical information. For more details on CBFs refer (Ames et al., 2019). Additionally safety in the learning process may also be ensured by incorporating physical constraints into the RL loss (Chen, Liu, Wang, & Kamwa, 2022; Li et al., 2021).
3) *Physics parameters, primitives and physical variables (PPV):* Physics values extracted/ derived from the environment or system has been directly used by RL agents in form of physics parameters (Siekmann et al., 2021), dynamic movement (physics) primitives (Bahl, Gupta, & Pathak, 2021), physical state (Jurj et al., 2021) and physical target (Li et al., 2023). For example in Li et al. (2023), the reward is created to meet two physical objectives/ targets: operation cost and self-energy sustainability. In an adaptive cruise control problem (Jurj et al., 2021), authors use desired physical parameters e.g. velocity and acceleration constraints and minimum jam avoiding distance, as a state space input.
4) *Offline data and representation (ODR):* For the improvement simulator based training, especially during sim-to-real transfer, non-task-specific-policy data collected from real robot has been used to train RL agents in offline setting along with simulators (Golemo et al., 2018) and as hardware data to seed simulators (Lowrey et al., 2018).

Another popular way of extracting physics information from environment is learning physically relevant low dimensional representation from observations (Cao et al., 2023a; Gokhale et al., 2022). For example, in Gokhale et al. (2022), PINN is used to extract physically relevant information about the hidden state of the system, which is further used to learn a Q-function for policy optimization.

5) *Physics simulator and model (PS):* Simulators provide a easy way of experimenting with RL algorithms without exposing the agent e.g. a robot to the wear and tear of the real environment.

   Apart from serving as test-beds for RL algorithms, simulators are also used alongside RL algorithms to impart physical correctness or physics awareness in the data or training process. For example in order to improve motion capture and imitation of given motion clips, Chentanez, Müller, Macklin, Makoviychuk, and Jeschke (2018) have used rigid body physics simulations to solve the rigid body poses closely following the motion capture clip. In Garcia-Hernando, Johns, and Kim (2020), using a physics simulator, a residual agent is able to learn how to improve user input in order to achieve a task while staying true to the original input and expert-recorded trajectories.

   In the MBRL setting the system model can be: 1) completely known, 2) partially known or 3) completely unknown. RL algorithms typically addresses the last two types, since it deals with environments whose dynamics is complex and difficult to ascertain through classical approaches. In such cases a DNN based data-driven approach is generally utilized to learn the system model completely or enrich the existing partial or basic model of the environment. In Han et al. (2022) a data driven surrogate traffic flow model is learned that generates synthetic data. This data is later used by the agent in an offline learning process, followed by an online control process. In Ramesh and Ravindran (2023) learns environment and reward models by using Lagrangian NNs (Cranmer et al., 2020). LNNs are models that are able to Lagrangian functions straight from data gathered from agent-environment interactions.

6) *Physical properties (PPR):* Fundamental knowledge regarding the physical structure or properties pertaining to a system has been used in a number of works. For example system morphology, system symmetry (Huang et al., 2023)

3.3.2. PIRL methods: Physics prior augmentations to RL

PIRL methods highlights and discusses about the different components of the typical RL paradigm e.g. state space, action space, reward function and agent networks (policy and value function N/W), that has been directly modified/ augmented through the incorporation of physics information.

1) *State design:* This category is concerned with the observed state space of the environment or model. The PIRL approaches, typically modifies or expands the state representation in order to make it more instructive. Works include state fusion using additional information from environment (Jurj et al., 2021) and other agents (Shi et al., 2023), state as extracted features from robust representation (Cao et al., 2023a), learned surrogate model generated data as state (Han et al., 2022) and state constraints (Zhang, Peng, Pan, Xu, & Xie, 2022).
2) *Action regulation:* This pertains to modifying the action value, which is often achieved through PIRL approaches that impose constraints on the action value to ensure safety protocols are implemented (Cheng et al., 2019; Li et al., 2021).
3) *Reward design:* It concerns approaches that induce physics information through effective reward design or augmentation of existing reward functions with bonuses or penalties (Dang & Ishii, 2022; Luo, Hachiuma, Yuan, Iwase, & Kitani, 2020).
4) *Augment policy or value N/W:* These PIRL approaches incorporate physics principles via methods like, adjusting the update rules and losses of the policy (Bahl et al., 2020; Margolis et al., 2021), value functions (Mukherjee & Liu, 2023; Park et al., 2023) and making direct changes to their underlying network structure (Cao et al., 2023b). Works with novel physics based losses (Mora, Peychev, Ha, Vechev, & Coros, 2021; Xu et al., 2022) and constraints for policy or value function learning (Gao et al., 2022) are also included.
5) *Augment simulator or model:* This category encompasses those works that develops improved simulators through incorporation of underlying physics knowledge thereby allowing for more accurate simulation of real-world environments. Works include physics based augmentation of DNN based learnable models for accurate system model learning (Lee et al., 2020; Ramesh & Ravindran, 2023), improved simulators for sim-to-real transfer (Golemo et al., 2018; Lowrey et al., 2018) and physics informed learning for partially known environment model (Liu & Wang, 2021).

3.3.3. RL pipeline

As illustrated in Fig. 6, a typical RL pipeline can be represented into four functional stages namely, the problem representation, learning strategy, network design, training and trained policy deployment. These stages are elaborated as follows:

1. *Problem representation:* In this stage, a real-world problem is modeled as a Markov Decision Process (MDP) and thereby described using formal RL terms. The main challenge is to choose the right observation vector, define the reward function, and specify the action space for the RL agent so that it can perform the specified task properly.
2. *Learning Strategy:* In this stage, the decisions are made regarding the type of agent-environment interaction e.g. in terms of environment model use, learning architecture and the choice of RL algorithm.
3. *Network design:* Here the finer details of the learning framework are decided and customized where needed. Decisions are made

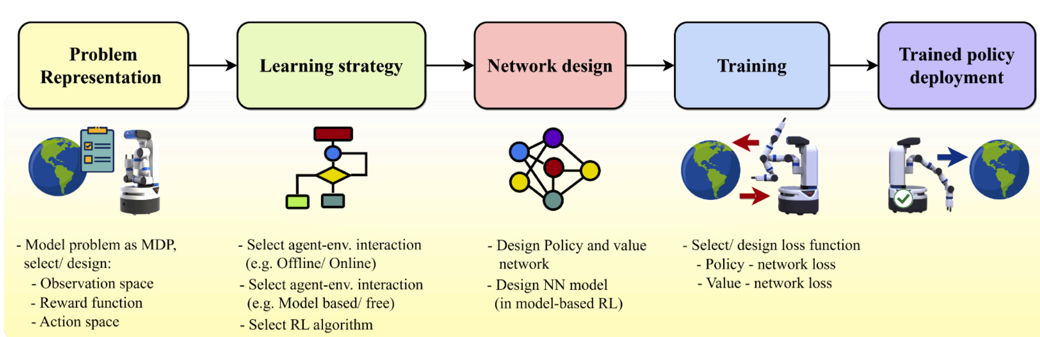


Fig. 6. Deep Reinforcement Learning Pipeline. Here the problem is first modeled as a MDP, clearly defining the state, action and reward spaces. Followed by selecting the RL algorithm as Learning strategy and then selecting/ designing the policy and/or value networks in network design stage. Finally the agent is trained using default/ custom loss function in training stage and finally deployed.

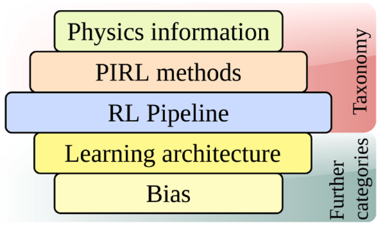


Fig. 7. PIRL taxonomy and further categories. Physics information (types), the RL methods that incorporate them and the underlying RL pipeline constitutes the PIRL Taxonomy, see Fig. 9. bias (sec. 3.4.1) and Learning architecture (sec. 3.4.2) are two additional categories which has been introduced to better explain the implementation of PIRL.

regarding the type of constituent units (e.g. layer types, network depth etc.) of underlying Policy and value function networks.

4. *Training:* The policy and allied networks are trained in this stage. It also represents training augmentation approaches like Sim-to-real, that helps is reducing discrepancy between simulated and real worlds.
5. *Trained policy deployment:* At this stage the policy is completely trained and is deployed for solving the concerned task.

3.4. Further categorization

In this section we introduce two additional categorizations: Bias and Learning architecture. These categories are not part of the taxonomy that we have discussed in the previous section, see Fig. 7. They provide an additional perspective to the PIRL approaches presented here.

3.4.1. Bias (in integration of physics in RL)

PI approaches in ML paradigm, mentions of different kind of biases or categories of methods of physics incorporation in ML models (refer to Sec. 2, for a detailed discussion on biases.) In order to relate to that existing taxonomy used in PIML methods, in Table 2 and Table 3, we include corresponding bias categories to each of the PIRL entries.

3.4.2. Learning architecture

We also categorize PIRL algorithms based on the alterations that they introduce to the conventional RL learning architecture to incorporate physics information/ priors. As listed and discussed below they help us understand the PIRL methods from an architectural point of view. In the literature review section we use the aid of such learning architecture categories to group and discuss the PIRL methods.

1) *Safety filter:* This category includes approaches that has a PI based module which regulates the agent's exploration ensuring safety constraints, for reference see Fig. 8(a). In this typical architecture the safety-filter module takes action a_t from RL agent π_φ, and state information (s_t) and refines the action, giving $\tilde{a}_t$.
2) *PI reward:* This category includes approaches where physics information is used to modify the reward function, see Fig. 8(b) for reference. Here the PI-reward module augments agent's extrinsic reward (r_t) with a physics information based intrinsic component, giving $\tilde{r}_t$.
3) *Residual learning:* Residual RL is an architecture which typically consists of two controllers: a human designed controller and a learned policy (Johannink et al., 2019). In PIRL setting the architecture consists of a physics informed controller π_ψ along with the data-driven DNN based policy π_φ, called residual RL agent, see Fig. 8(c).
4) *Physics embedded network:* In this category physics information e.g. system dynamics is directly incorporated in the policy or value function networks, see Fig. 8(d) for reference.
5) *Differentiable simulator:* Here the approaches have use differentiable physics simulators, which are non-conventional/ or adapted simulators and explicitly provides loss gradients of simulation outcome w.r.t. control action, see Fig. 8(e) for reference.
6) *Sim-to-real:* In Sim-to-real architecture, the agent is first trained on a simulator or source domain and is later transferred to a target domain for deployment. In certain cases the transfer is followed by fine-tuning at the target domain, see Fig. 8(f) for reference.
7) *Physics variable:* This architecture encompasses all those approaches where physical parameters, variables or primitives are

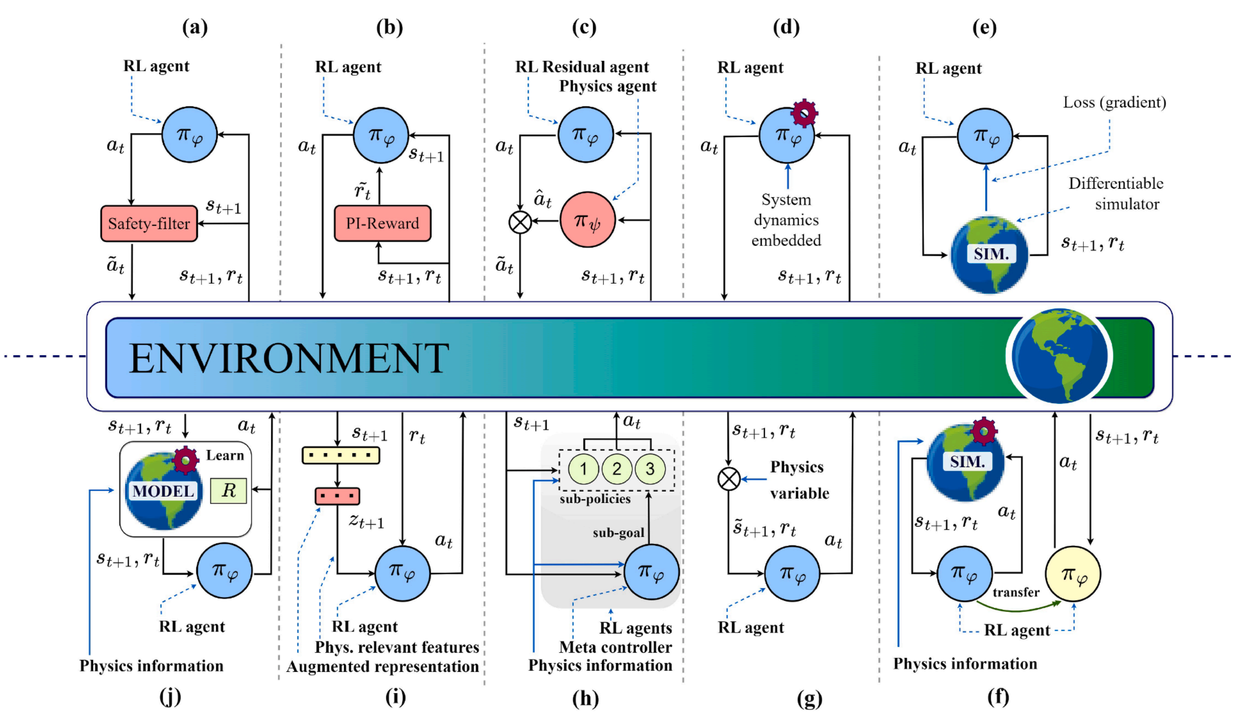


Fig. 8. Typical RL architectures with physics information incorporation (a) Safety filter (b) PI-Reward (c) Residual agent (d) Physics embedded network (e) Differentiable simulator (f) Sim-to-Real (g) Physics variable (h) Hierarchical RL (i) Data augmentation (j) PI model identification. To keep illustrations simple, we have not included ancillary networks e.g. value function networks above.

Table 2
Summary of PIRL Literature - Model Free.

Ref.	Year	Context/ Application	RL Algorithm	Learning arch.	Bias	Physics information	PIRL methods	RL pipeline
Chentanez et al. (2018)	2018	Motion capture	PPO	Physics reward	Learning	Physics simulator	Reward design	Problem representation
Peng, Abbeel, Levine, and Van de Panne (2018)	2018	Motion control	PPO (Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017)	Physics reward	Learning	Physics simulator	Reward design	Problem representation
Golemo et al. (2018)	2018	Policy optimization	PPO	Sim-to-Real	Observational	Offline data	Augment simulator	Training
Lowrey et al. (2018)	2018	Policy optimization	NPG (Williams, 1992) (C)[a]	Sim-to-Real	Observational	Offline data	Augment simulator	Training
Cho et al. (2019)	2019	Molecular structure optimization	DDPG	Physics reward	Learning	DFT (PS)	Reward design	Problem representation
							Augment simulator	Training
Li and Belta (2019)	2019	Safe exploration and control	PPO	Residual RL	Learning	CBF, CLF, FSA/TL (BPC)	Reward design	Problem representation
							Augment policy	Learning strategy
Bahl et al. (2020)	2020	Dynamic system control	PPO	Phy. embed. N/W	Inductive	DMP (PPV)	Augment policy	Network design
Garcia-Hernando et al. (2020)	2020	Dexterous manipulations	PPO	Residual RL	Observational	Physics simulator	Reward design	Problem representation
Luo et al. (2020)	2020	3D Ego pose estimation	PPO	Physics reward	Learning	Physics simulator	State, Reward design	Problem representation
Bahl et al. (2021)	2021	Dynamic system control	PPO	Hierarchical RL	Inductive	DMP (PPV)	Augment policy	Network design
Margolis et al. (2021)	2021	Dynamic system control	PPO	Hierarchical RL	Learning	WBIC (PPV)	Augment policy	Learning strategy
Alam et al. (2021)	2021	Manufacturing	SARSA (Sutton & Barto, 1998)	Sim-to-Real	Observational	Physics engine	Augment simulator	Training
Siekmann et al. (2021)	2021	Dynamic system control	PPO	Phy. variable	Learning	Physics parameters	Reward design	Problem representation
Li et al. (2021)	2021	Safe exploration and control	NFQ (Riedmiller, 2005)	Safety filter	Learning	Physical constraint	Action regulation	Problem representation
Jurj et al. (2021)	2021	Safe cruise control	SAC	Phy. variable	Observational	Physical state (PPV)	State design	Problem representation
Mora et al. (2021)	2021	Policy optimization	DPG (C)	Diff. Simulator	Learning	Physics simulator	Augment policy	Learning strategy
Radaideh et al. (2021)	2021	Optimization, nuclear engineering	DQN, PPO	Physics reward	Learning bias	Physical properties (PPR)	Reward design	Problem representation
Zhao and Liu (2021)	2021	Air-traffic control	PPO	Data augmentation	Observational	Representation (ODR)	State design	Problem representation
Wang (2022)	2022	Motion planner	PPO + AC (Konda & Tsitsiklis, 1999)	Safety filter	Learning	CBF (BPC)	Action regulation Reward design	Problem representation
Chen et al. (2022)	2022	Active voltage control	TD3 (C)	Safety filter	Learning	Physical constraints	Penalty function	Problem representation
							Action regulation	
Dang and Ishii (2022)	2022	Interfacial structure prediction	DDPG	Off-policy	Learning	Physics model	Reward design	Problem representation
Gao et al. (2022)	2022	Transient voltage control	DQN	PINN loss	Learning	PDE (DAE)	Augment policy	Learning strategy
Gokhale et al. (2022)	2022	Building control	Q-learning (C)	Data augment	Observational	Representation (ODR)	State design	Problem representation
Han et al. (2022)	2022	Traffic control	Q-Learning	Data augment	Observational	Physics model	State design	Problem representation
Martin and Schaub (2022)	2022	Safe exploration and control	SAC	Sim-to-Real	Observational	Physics model	Augment simulator	Training
Jiang, Fu, and Chen (2022)	2022	Dynamic system control	SAC (etc.)	Physics reward	Learning	Barrier function	Reward design	Problem representation
Xu et al. (2022)	2022	Policy Learning	Actor-critic (C)	Diff. Simulator	Learning	Physics simulator	Augment policy	Learning strategy
Cao et al. (2023c)	2023	Safe exploration and control	DDPG	Residual RL	Learning	Physics model	Reward design	Problem representation
							Action regulation	

continued

Table 2
(*continued*).

Ref.	Year	Context/ Application	RL Algorithm	Learning arch.	Bias	Physics information	PIRL methods	RL pipeline
Cao et al. (2023b)	2023	Safe exploration and control	DDPG	Residual RL	Inductive	Physics model	Reward design	Problem representation
							Action regulation	
					Inductive		N/W editing (Aug. pol.)	Network design
Cao et al. (2023a)	2023	Robust voltage control	SAC	Data augment	Observational	Representation (ODR)	State design	Problem representation
Chen, Liu, and Di (2023b)	2023	Mean field games	DDPG	Physics reward	Learning	Physics model	Reward design	Problem representation
Yang et al. (2023)	2023	Safe exploration and control	PPO (C)	Safety filter	Learning	NBC (BPC)	Augment policy	Training
Zhao et al. (2023)	2023	Power system stability enhancement	Custom	Safety filter	Learning	NBC (BPC)	Action regulation	Problem representation
Du et al. (2023)	2023	Safe exploration and control	AC (C)	Safety filter	Learning	CLBF (Dawson, Qin, Gao, & Fan, 2022; Romdlony & Jayawardhana, 2016) (BPC)	Augment value N/W	Training
Shi et al. (2023)	2023	Connected automated vehicles	DPPO	Physics variable	Observational	Physical state (PPV)	State design	Problem representation
					Learning		Reward design	
Korivand et al. (2023)	2023	Musculoskeletal simulation	SAC (C)	Physics variable	Learning	Physical value	Reward design	Problem representation
Li et al. (2023)	2023	Energy management	MADRL(C)	Physics variable	Learning	Physical target	Reward design	Problem representation
Mukherjee and Liu (2023)	2023	Policy optimization	PPO	Phy. embed N/W	Inductive	PDE (DAE)	Augment value N/W	Network design
Yousif et al. (2023)	2023	Flow field reconstruction	A3C	Physics reward	Learning	Physical constraints	Reward design	Problem representation
Park et al. (2023)	2023	Freeform nanophotonic devices	$\epsilon-$ greedy Q	Phy. embed N/W	Inductive	ABM	Augment value N/W	Network design
Rodwell and Tallapragada (2023)	2023	Dynamic system control	DPG	Curriculum learning	Learning	Physics model	Augment simulator	Training
She et al. (2023)	2023	Energy management	TD3	Sim-to-Real	Observational	Physics model	Augment simulator	Learning strategy
Yin et al. (2023)	2023	Robot wireless navigation	PPO	Physics reward	Learning	Physical value	Reward design	Problem representation

[a] C represents custom versions of the adjacent conventional algorithms.

introduced to augment components (e.g. states and reward) of the RL framework. For reference see Fig. 8(g).

8) *Hierarchical RL:* This category includes hierarchical and curriculum learning based approaches, Fig. 8(h) for reference. In a hierarchical RL (HRL) setting a long horizon decision making task is broken into simpler sub-tasks autonomously. In curriculum learning a complex task is solved by learning to solve a series of increasingly difficult tasks. In both HRL and CRL physics is typically incorporated into all the policy (including meta and sub-policies) and value networks. Approaches here are mostly extensions of physics-embedded networks (Fig. 8(d)), as used in non-HRL/ CRL settings.
9) *Data augmentation:* This category includes approaches where the input state is replaced with a different or augmented form of it, e.g. low dimensional representation so as to derive special and physically relevant features out of it. See Fig. 8(i) for reference. In this typical architecture, the state vector s_{t+1} is transformed into an augmented representation z_{t+1}. Physically relevant features are then extracted from it and used by the RL agent (π_{φ}).
10) *PI Model Identification:* This architecture represents those PIRL approaches, especially in data-driven MBRL setting where physics information is directly incorporated into the model identification process. For reference see Fig. 8(j).

4. PIRL: Review and analysis

In this section we provide a indepth review of latest works in PIRL, followed by a review of the popular datasets. We also include an analysis of the algorithms and their derivatives, and discuss crucial insights.

4.1. Algorithmic review

We provide a detailed overview of the PIRL approaches as identified by our literature review in Table 2 and Table 3. We have structured our discussion according to the methods of the introduced taxonomy (see 3.3) since they form a bridge between the physics information sources and practical applications. We also use learning architecture categories as introduced in 3.4.2, to better explain the PIRL methods.

4.1.1. State design

Vehicular traffic control applications have used physics priors to design the state representations. While controlling connected automated vehicles (CAVs), Shi et al. (2023) proposed the use of surrounding information from downstream vehicles and roadside geometry, by embedding them in the state representation, see Fig. 10. The physics-informed state fusion approach integrates received information as DRL

Table 3
Summary of PIRL literature - Model based.

Ref.	Year	Context/ Application	Algorithm	Learning arch.	Bias	Physics information	PIRL method	RL pipeline
Xie et al. (2016)	2016	Exploration and control	-	Model learning	Observational	Sys. morphology (PPR)	Augment model	Learning strategy
Sanchez-Gonzalez et al. (2018)	2018	Dynamic system control	-	Model learning	Inductive	Physics model	Augment model	Learning strategy
Ohnishi et al. (2019)	2019	Safe navigation	-	Safety filter	Learning	CBC (BPC)	Action regulation	Problem representation
Cheng et al. (2019)	2019	Safe exploration and control	TRPO, DDPG	Residual RL	Learning	CBF (BPC)	Action regulation	Problem representation
Veerapaneni et al. (2020)	2020	Control (visual RL)	-	Model learning	Observational	Entity abstraction (ODR)	Augment model	Learning strategy
Lee et al. (2020)	2020	Dynamic system control	-	Model learning	Observational	Context encoding (ODR)	Augment model	Learning strategy
Choi et al. (2020)	2020	Safe exploration and control	DDPG (Silver et al., 2014)	Safety filter	Learning	CBF, CLF, QP (BPC)	augment policy	Learning strategy
Liu and Wang (2021)	2021	Dynamic system control	Dyna + TD3(C)[a]	Model identification	Learning	PDE/ ODE, BC (DAE)	Augment model	Learning strategy
Duan et al. (2021)	2021	Dynamic system control	PPO	Residual-RL	Learning	Physics model	Action regulation	Problem representation
Cai et al. (2021)	2021	Multi agent collision avoidance	MADDPG (C)	Safety filter	Learning	CBF (BPC)	Action regulation	Problem representation
Lv et al. (2022)	2022	Dynamic system control	TD3(C)	Sim-to-Real	Learning	Physics simulator	Augment policy	Learning strategy
Udatha, Lyu, and Dolan (2022)	2022	Traffic control	AC (Ma et al., 2021)	Safety filter	Learning	CBF (BPC)	Augment model	Learning strategy
Zhao et al. (2022)	2022	Safe exploration and control	DDPG	Safety filter	Learning	CBC (BPC)	Augment policy	Learning strategy
Zhang et al. (2022)	2022	Distributed MPC	AC (Jiang, Fan, Gao, Chai, & Lewis, 2020)	Safety filter	Learning	CBF (BPC)	State design	Problem representation
Ramesh and Ravindran (2023)	2023	Dynamic system control	Dreamer (Hafner, Lillicrap, Ba, & Norouzi, 2019)	Phy. embed. N/W	Inductive	Physics model	Augment model	Network design
Cohen and Belta (2023)	2023	Safe exploration and control	-	Safety filter	Learning	CBF (BPC)	Augment model	Learning strategy
Huang et al. (2023)	2023	Attitude control	-	Phy. embed N/W	Inductive	System symmetry (PPR)	Augment model	Network design
Wang, Cao, Zhou, Wen, and Tan (2023)	2023	Data center cooling	SAC	Model identification	Learning	Physics laws (PPR)	Augment model	Learning strategy
Yu, Zhang, and Song (2023)	2023	Cooling system control	DDPG	Residual RL	Learning	CBF (BPC)	Action regulation	Problem representation

[a] C represents custom versions of the adjacent conventional algorithms.

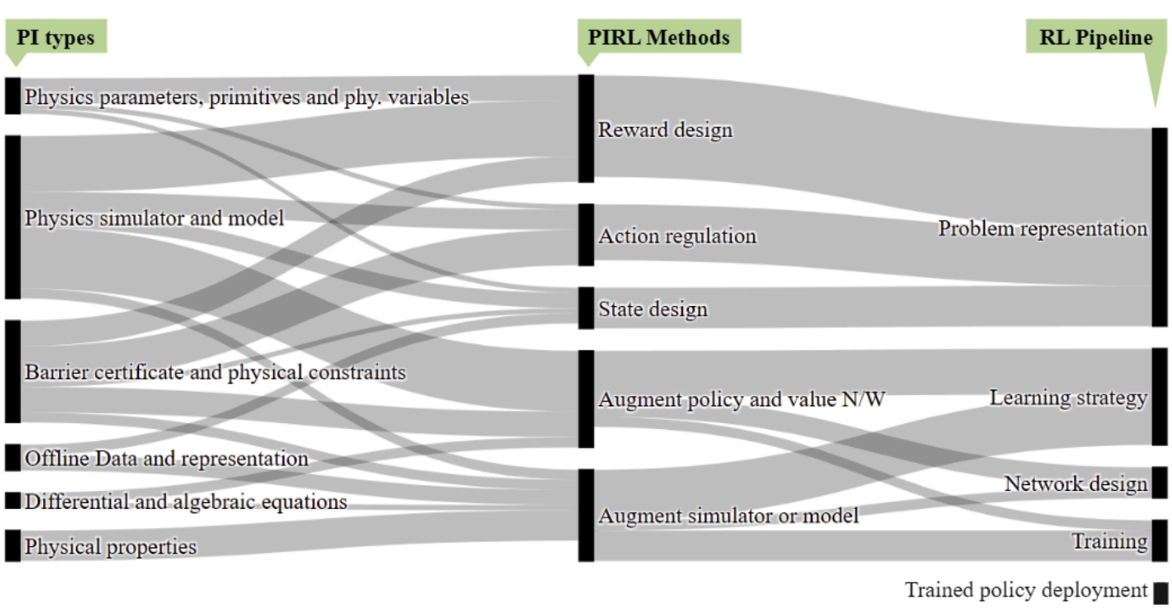


Fig. 9. PIRL Taxonomy, the diagram connects PI types with PIRL methods and then to the RL pipeline backbone. The connection thickness represents the quantity of work done which corresponds to those components/ categories.

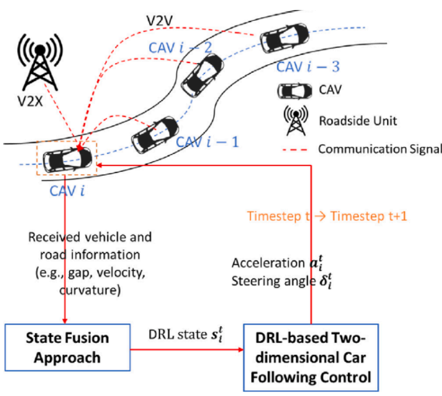


Fig. 10. Example of state design, through physics incorporation. Distributed control framework for connected automated vehicles (Shi et al., 2023). Here information from downstream vehicles and roadway geometry information are incorporated as physics prior knowledge through state fusion.

state (input features) i.e., for the i^{th} CAV, DRL state is given as $s_i^t = [e_i^t, \phi_i^t, \delta q_i^{-t}, \delta d_i^{-t}, k_i^t]$, which are deviation values, (from left): lateral, angular, weighed equilibrium spacing and speed, and road curvature information.

Jurj et al. (2021) makes use of physical information like jam-avoiding distance to train RL agent, in order to improve collision avoidance of vehicles with adaptive cruise control. In ramp metering control, Han et al. (2022) utilizes an offline-online policy training process, where the offline training data consists of historical data and synthetic data generated from a physical traffic flow model.

In Cao et al. (2023a) a physics informed graphical representation-enabled, global graph attention (GGAT) network is trained to model power flow calculation process. Informative features are then extracted from the GGAT layer (as representation N/W) and transferred used in the policy training process. While (Gokhale et al., 2022), uses PINNs based on thermal dynamics of buildings for learning better heating control strategies. Dealing with aircraft conflict resolution problem, Zhao and Liu (2021) composed intruder's information e.g. speed and heading angle into an image state representation. This image now constitutes of the physics prior and serves as the input feature for RL based learning. In Zhang et al. (2022), the authors proposed a safe reinforcement learning algorithm using barrier functions for distributed MPC nonlinear multi-robot systems, with state constraints. Ohnishi et al. (2019), incorporates trained model alongside control barrier certificates, which restrict policies and prohibits exploration of the RL agent into certain undesirable sections of the state space. In case of a safety breach due to non-stationarity, the Lyapunov stability conditions ensures the re-establishment of safety.

4.1.2. Action regulation

Many safety critical applications have used physics based constraints and other information in action regulation. These kind of approaches can be categorized under shielded RL/ safety filter, where a type of safety shield or barrier function is employed to check the actions.

For safe power system control Zhao et al. (2023) proposes a framework for learning a stabilizing controller that satisfies predefined safety regions, see Fig. 11. Combining a model-free controller and a barrier-certification system, using a NN based barrier function, i.e., neural barrier certificate (NBC). Given a training set they learn a NBC $B_\epsilon(x)$ and filtered (regulated) control action $\mathcal{F}_u^\psi$, jointly holding the following condition

$$(\forall x \in \mathcal{S}_0, B_\epsilon(x) \le 0) \wedge (\forall x \in \mathcal{S}_u, B_\epsilon(x) > 0)$$
$$\wedge (\forall x \in x | B_\epsilon(x) = 0, \mathcal{L}_{f(x,u_{RL})} B_\epsilon(x) < 0)$$

where $\mathcal{L}_{f(x,u_{RL})} B_\epsilon(x)$ is the Lie derivative of $B_\epsilon(x)$, and ϕ, ϵ are NN parameters. $\mathcal{S}_0, \mathcal{S}_u$ are set of initial states and unsafe states respectively.

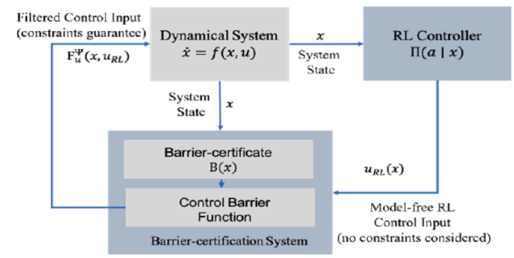


Fig. 11. Example of action regulation, using physics priors. In Zhao et al. (2023), a barrier certification system receives RL control policy generated control actions and refines them sequentially using a barrier certificate to satisfy operational constraints.

Cheng et al. (2019) introduces a hybrid approach of MFRL and MBRL using CBF, with provision of online learning of unknown system dynamics. It assumes availability of a set of safe states. In a MARL setting, Cai et al. (2021) introduced cooperative and non-cooperative CBFs in a collision-avoid problem, which includes both cooperative agents and obstacles. Also in MARL setting, Chen et al. (2022) proposed efficient active voltage controller of photovoltaics (PVs) enabled with shielding mechanism. Which ensures safe actions of battery energy storage systems (BESSs) during training. Yu et al. (2023) deals with controlling a district cooling system (DCS), with complex thermal dynamic model and uncertainties from regulation signals and cooling demands. The proposed safe controller a hybrid of barrier function and DRL and helps avoid unsafe explorations and improves training efficiency. Li et al. (2021) proposed a safe RL framework for adaptive cruise control, based on a safety-supervision module. The authors used the underlying system dynamics and exclusion-zone requirement to construct a safety set, for constraining the learning exploration.

In a highway motion planning setting for autonomous vehicles (Wang, 2022) proposed a CBF-DRL hybrid approach. Certain works like (Cao et al., 2023c) and (Cao et al., 2023b) have introduced multiple physics based artifacts to ensure safe learning in autonomous agents. Both of them used residual control based architecture merging physical model and data driven control. Additionally it also leverages physics model guided reward. Cao et al. (2023b) extends the work by Cao et al. (2023c) and introduces physics model guided policy and value network editing in addition to the physics based reward. In Duan et al. (2021), the authors integrate learning a task space policy with a model based inverse dynamics controller, which translates task space actions into joint-level controls. This enables the RL policy to learn actions in task space.

4.1.3. Reward design

In sim-to-real setting (Siekmann et al., 2021) proposed a reward specification framework based on composing probabilistic periodic costs on basic forces and velocities, see Fig. 12. The framework defines a parametric reward function for common robotic (bipedal) gaits. Dealing with periodic robot behavior, the absolute time reward function is here defined in terms of a cycle time variable ϕ (which cycles over time period of $[0, 1]$, as $R(s, \phi)$. The updated reward function as given below, is defined as a biased sum of n reward components $R_i(s, \phi)$, each capturing a desired robot gait characteristic.

$$R(s, \phi) = \beta + \Sigma R_i(s, \phi), \text{ where}$$
$$R_i(s, \phi) = c_i \times I_i(\phi) \times q_i(s)$$

each $R_i(s, \phi)$ is a product of phase-coefficient c_i, phase indicator $I_i(\phi)$ and phase reward measurement $q_i(s)$.

In Chen et al. (2023b), the authors introduced a RL-PIDL hybrid framework, to learn MFGs, which generalize well and manage can complex multi-agent systems applications. The physics based reward component (= evolution of population density/ mean-field state) is approximated using PINN. To better mimic natural human locomotion (Korivand et al., 2023), designed reward function based on physical

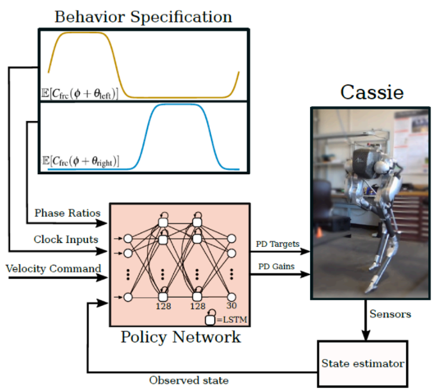


Fig. 12. Example of physics incorporation in reward design. In Siekmann et al. (2021) a reward function design framework was introduced, that describe robot gaits as a periodic phase sequence such that each of which rewards or penalizes a particular physical system measurement.

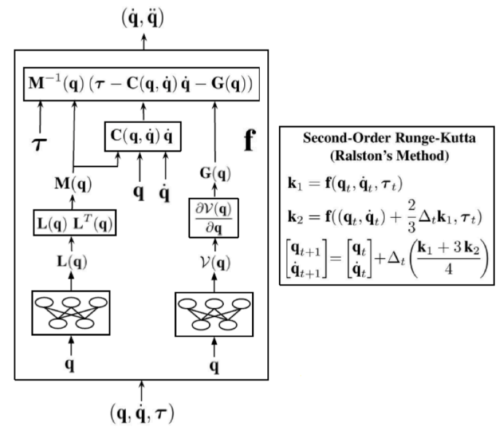


Fig. 13. Example, augmentation of learnable model using physics information. The figure shows system dynamics learning network structured using a LNN (Ramesh & Ravindran, 2023) and next state calculations using Ralston's method. Here the PINN (LNN) based dynamics model and reward model , are learned via data-driven method.

and experimental information: trajectory optimization rewards, and bio-inspired rewards. In a similar task of imitation of human motion but from motion clip, Chentanez et al. (2018) proposes a physics-based controller using DRL. A rigid body physics simulator is used to solve rigid body poses that closely follows the motion capture (mocap) clip frames. In a similar work (Peng et al., 2018), a data driven RL framework was introduced for training control policies for simulated characters. Refernce motions are used to define imitation reward and the task goal defines task specific reward.

Li et al. (2023) leverages a federated MADRL approach for energy management in multi-microgrid settings. The reward is designed to satisfy two physical targets: operation cost and self energy sufficiency. Yousif et al. (2023) proposed a DRL based method for reconstruction of flow fields from noisy data. Physical constraints like momentum equation, pressure Poisson equation and boundary conditions are used for designing the reward function. Yin et al. (2023) proposed physics based reward shaping for wireless navigation applications. They used a cost function augmented with physically motivated costs like costs for link-state monotonicity, for angle of arrival direction following, and for SNR increasing. In single molecule 3D structure optimization problem, Cho et al. (2019) used physics based DFT calculation is used as reward function, for physically correct structural prediction. In Li and Belta (2019), the authors used temporal logic through a finite state automata (FSA), control Lyapunov and barrier function for ensuring effective and safe RL in complex environments.The FSA simultaneously provides rewards, objectives and safety constraints to the framework components.

Addressing the problem of dexterous manipulation of objects in virtual environments, Garcia-Hernando et al. (2020) trained the agent in a residual setting using hybrid model-free RL-IL approach. Using a physics simulator and a pose estimation reward the agent learns to refine the user input to achieve a task while keeping the motion close to the input and the expert demonstrations. Luo et al. (2020) tackles physically valid 3D pose estimation from egocentric video. The authors utilized a combination of kinematics and dynamics approach, whereby the residual of the action against a learned kinematics model is outputted by the dynamics-based model. In Jiang et al. (2022),the authors proposed inclusion of physics based intrinsic reward for improved policy optimization of RL algorithms.

In the context of predicting interfacial area in two-phase flow, Dang and Ishii (2022) proposed. The two-phase flow physics information is infused into the underlying MDP framework, which is then uses RL strategies to describe behavior of flow dynamics. The work introduces multiple rewards based on physical interfacial area transport models, other physical parameters and data. In a work concerning optimization of nuclear fuel assembly (Radaideh et al., 2021), the authors introduce a reward shaping approach in RL optimization, which is based on physical tactics used by fuel designers. These tactics include moving fuel rods in assembly to meet certain constraints and objectives.

A number of works have used physics through multiple PIRL methods. Apart from reward design they have infused physics information through state design (Shi et al., 2023) and action regulation (Cao et al., 2023b,c; Wang, 2022). They have been discussed in previous sections and hence not repeated.

4.1.4. Augment simulator or model

In MBRL setting, using structure of underlying physics, and building upon Lagrangian neural network (LLN) (Cranmer et al., 2020), Ramesh and Ravindran (2023) learned the system model via data-driven approach, see Fig. 13. Concerning systems obeying Lagrangian mechanics, the state consists of generalized coordinates q and velocities $\dot{q}$. Lagrangian, which is a scalar is defined as

$$\mathcal{L}(q, \dot{q}, t) = \mathcal{T}(q, \dot{q}) - \mathcal{V}(q)$$

where $\mathcal{T}(q, \dot{q})$ is kinetic energy and $\mathcal{V}(q)$ is the potential energy. And so the Lagrangian equation of motion can be written as

$$\tau = M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q), \text{where}$$

$$\ddot{q} = M^{-1}(q)(\tau - C(q, \dot{q})\dot{q} - G(q))$$

where $C(q, \dot{q})\dot{q}$ is Coriolis term, $G(q)$ is gravitational term and τ is motor torque. In the NN implementation, separate networks are used for learning $\mathcal{V}(q)$ and $L(q)$, leveraging which the acceleration ($\ddot{q}$) quantity is generated. The output state derivative ($\dot{q}, \ddot{q}$) is then integrated using 2^{nd}-order Runge-Kutta to compute next state.

Concerning a sim-to-real setting, in Golemo et al. (2018) authors train a recurrent neural network on the differences between robotic trajectories in simulated and actual environments. This model is further used to improve the simulator. For improved transfer to real environment, Lowrey et al. (2018) collected hardware data (positions and calculated system velocities) to seed the simulator, for training control policies. Alam et al. (2021) proposes a framework for autonomous manufacturing of acoustic meta-material, while leveraging physics informed RL and transfer learning. A physics guided simulation engine is used to train the agent in source task and then fine-tuned in a data-driven fashion in the target task.

Martin and Schaub (2022) introduced a PINN based gravity model for training of dynamically informed RL agents. Rodwell and Tallapragada (2023) uses surrogate models that capture primary physics of the system, as a starting point of training DRL agent. In a curriculum learning setting, they train an agent to first track limit cycles in a velocity space for a representative non-holonomic system and then further trained on a small simulation dataset. Xie et al. (2016) combines linear dynamic models of physical systems with optimism driven exploration. Here the features for the linear models obtained from robot morphology and the exploration is done using MPC.

A number of works introduced novel models are better representations of real world physics and serves as better simulators and ensures effective sim to real transfers. Sanchez-Gonzalez et al. (2018) introduced learnable physics models which supports accurate predictions and efficient generalization across distinct physical systems. Concerning dynamic control with partially known underlying physics (governing laws), Liu and Wang (2021) proposed a physics informed learning architecture, for environment model. ODEs and PDEs serves as the primary source of physics for these models. Veerapaneni et al. (2020) uses entity abstraction to integrate graphical models, symbolic computation and NNs in a MBRL agent. The framework presents object-centric perception, prediction and planning which helps agents to generalize to physical tasks not encountered before. Lee et al. (2020) proposes a context aware dynamics model which is adaptable to change in dynamics. They break the problem of learning the environment dynamics model into two stages: learning context latent vector and predicting next state conditioned on it.

In micro-grid power control problem, She et al. (2023) combines model-based analytical proof and reinforcement learning. Here model-based derivations are used to narrow the learning space of the RL agen, reducing training complexity significantly. In visual model based RL, Veerapaneni et al. (2020) models a scene in terms of entities and their local interactions, thus better generalizing to physical task the learner has not seen before. Similar to learning entity abstractions, in Lee et al. (2020) the authors tackles the challenge of learning a generalizable global model through: learning context latent vector, capturing local dynamics and predicting next state conditioned on the encoded vector. Addressing dynamic control problem in MBRL setting, Liu and Wang (2021) leveraged physical laws (in form of canonical ODE/ PDE) and environmental constraints to mitigate model bias issue and sample inefficiency. In autonomous driving safe ramp merging problem, Udatha et al. (2022) embedded probabilistic CBF in RL policy in order to learn safe policies, that also optimize the performance of the vehicle. Typically CBFs need good approximation of car's model. Here the probabilistic CBF is used as an estimate of the model uncertainty.

Cho et al. (2019) incorporates physics through reward design as well as through simulator augmentation, and has been discussed in previous section.

4.1.5. Augment policy and/or value N/W

In Bahl et al. (2020), proposes Neural Dynamic Policies (NDP) where they incorporate dynamical system as a differentiable layer in the policy network, see Fig. 14. In NDP, a NN Φ takes an input state (s_t) and predicts parameters of the dynamical system (i.e., (w, g)). Which are then used to solve second-order differential equation $\ddot{y} = \alpha(\beta(g-y) - \dot{(y)}) + f(x)$, to obtain system states $(y, \dot{y}, \ddot{y})$., which represents the behavior of the dynamic system, given a state goal g. Here α, β are global parameters allowing critical damping of system and f is a non-linear forcing function which primarily captures the shape of trajectory. Depending on robot's coordinate system an inverse controller may also be used to convert y to a, i.e., $a = \Omega(y, \dot{y}, \ddot{y})$. The NDPs thus can be defined as

$$\pi(a|s;\theta) \triangleq \Omega(DE(\Phi(s;\theta))), \text{where}$$
$$DE(w, g) \rightarrow \{y, \dot{y}, \ddot{y}\}$$

here $DE(w, g)$ represents solution of the differential equation.

Extending this work to hierarchical deep policy learning framework, Bahl et al. (2021) introduced H-NDP which forms a curriculum by learning local dynamical system-based policies on small state-space region and then refines them into global dynamical system based policy. Given the accurate dynamics and constraint of the system (Zhao et al., 2022) introduces control barrier certificates into actor-critic RL framework, for learning safe policies in dynamical systems. Margolis et al. (2021) proposes a method for generating highly agile and visually guided locomotion behaviors. They leverage MFRL while using model based optimization of ground reaction forces, as a behavior regularizer.

In Du et al. (2023) proposes an approach of safe exploration using CLBF without explicitly employing any dynamic model. The approach approximate the RL critic as a CLBF, from data samples and parameterized with DNNs. Both the actor and critic satisfies reachability and safety guarantees. Mukherjee and Liu (2023) combines PINN with RL, where the value function is treated as a PINN to solve Hamilton-Jacobi-Bellman (HJB) PDE. It enables the RL algorithm to exploit the physics of environment aswell as optimal control to improve learning and convergence.

Park et al. (2023) proposes an optimization method for freeform nanophotonic devices, by combining adjoint based methods (ABM) and RL. In this work the value network is initialized with adjoint gradient predicting network during initialization of RL process. Cao et al. (2023b) have used physics model to influence reward function, as well as edit policy and value networks as necessary. The work has been mentioned before in reward design.

To improve policy optimization, Mora et al. (2021) used differentiable simulators to directly compute the analytic gradient of the policy's value function w.r.t. the actions generated by it. This gradient information is used to monotonically improve the policy's value function. Gao et al. (2022) proposes a transient voltage control approach, by integrating physical and data-driven models of power system. They also uses the constraint of the physical model on the data-driven model to speed up convergence. A PINN trained using PDE of transient process acts as the physical model and contributes directly to the loss of the RL algorithm.

Xu et al. (2022) presents an efficient differentiable simulator (DS) with a new policy training algorithm which can effectively leverage simulation gradients. The learning algorithm alleviates issues inherent in DS while allowing many physical environments to be run in parallel. Chen et al. (2022) incorporates physics through action regulation and penalty signal to agent, and has been discussed in previous section.

In MBRL setting, Lv et al. (2022) leverage differentiable physics-based simulation and differentiable rendering. By comparing raw observations between simulated and real world, the initial learned system model is continually updated, producing a more physically consistent model. In data center (DC) cooling control application, Wang et al. (2023) proposed a lifelong-RL approach under evolving DC environment. It leverages physical laws of thermodynamics and the system and models the DC thermal transition and power usage through data collected online. Utilizing learned state transition and reward models it accelerates online adaptation.

Working with a nominal system model, Choi et al. (2020) presented an RL framework where the agent learns model uncertainty in multiple general dynamic constraints, e.g. CLF and CBF, through data-driven training. A quadratic program then solves for the control that satisfies the safety constraints under learned model uncertainty.

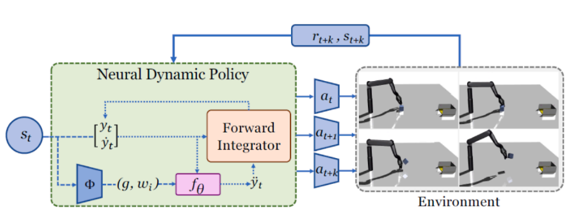


Fig. 14. Example, augmentation of policy using physics information. In Bahl et al. (2020), given an observation s_t from the environment, a neural dynamic policy generates w i.e., the weights of basis function and g which is a goal for the robot, for a function f_θ. This function is then used by an open loop controller to generate a set of actions from the robot to execute in the environment and collect next states and rewards to train the policy.

4.2. Review of simulation/ Evaluation benchmarks

In Table 4, we present the different training and evaluation benchmarks that has been used in the reviewed PIRL literature. We list the important insights from the table:

1. A majority works dealing with dynamic control have used OpenAI Gym (Xie et al., 2016), Safe Gym (Yang et al., 2023), MuJoCo (Veerapaneni et al., 2020; Zhou, Pinto, & Gupta, 2018), Pybullet (Du et al., 2023) and Deep mind control suite environments (Ramesh & Ravindran, 2023; Sanchez-Gonzalez et al., 2018), which are standard benchmarks in RL . Works dealing specifically with traffic management have used platforms like SUMO (Wang, 2022) and CARLA (Udatha et al., 2022).
2. Works dealing with power and voltage management problems have used IEEE distribution system benchmarks (Chen et al., 2022; Gao et al., 2022) to evaluate proposed algorithms. Alternatively in some works MATLAB/ SIMULINK platform is also used for training or evaluating RL agents (She et al., 2023)
3. One crucial observation is that a huge number of work have used customized or adapted environments for training and evaluation and have not used conventional environments (Cohen & Belta, 2023; Li & Belta, 2019; Lutter, Silberbauer, Watson, & Peters, 2021).

4.3. Statistical insights and trends in PIRL

4.3.1. Statistical analysis of PIRL literature

Use of RL Algorithms: As is evident from Fig. 15 (a), PPO (Schulman et al., 2017) and its variants are the most preferred RL algorithm, followed by DDPG (Silver et al., 2014). Among the comparatively new algorithms SAC (Haarnoja, Zhou, Abbeel, & Levine, 2018) is preferred over TD3 (Fujimoto, Hoof, & Meger, 2018).

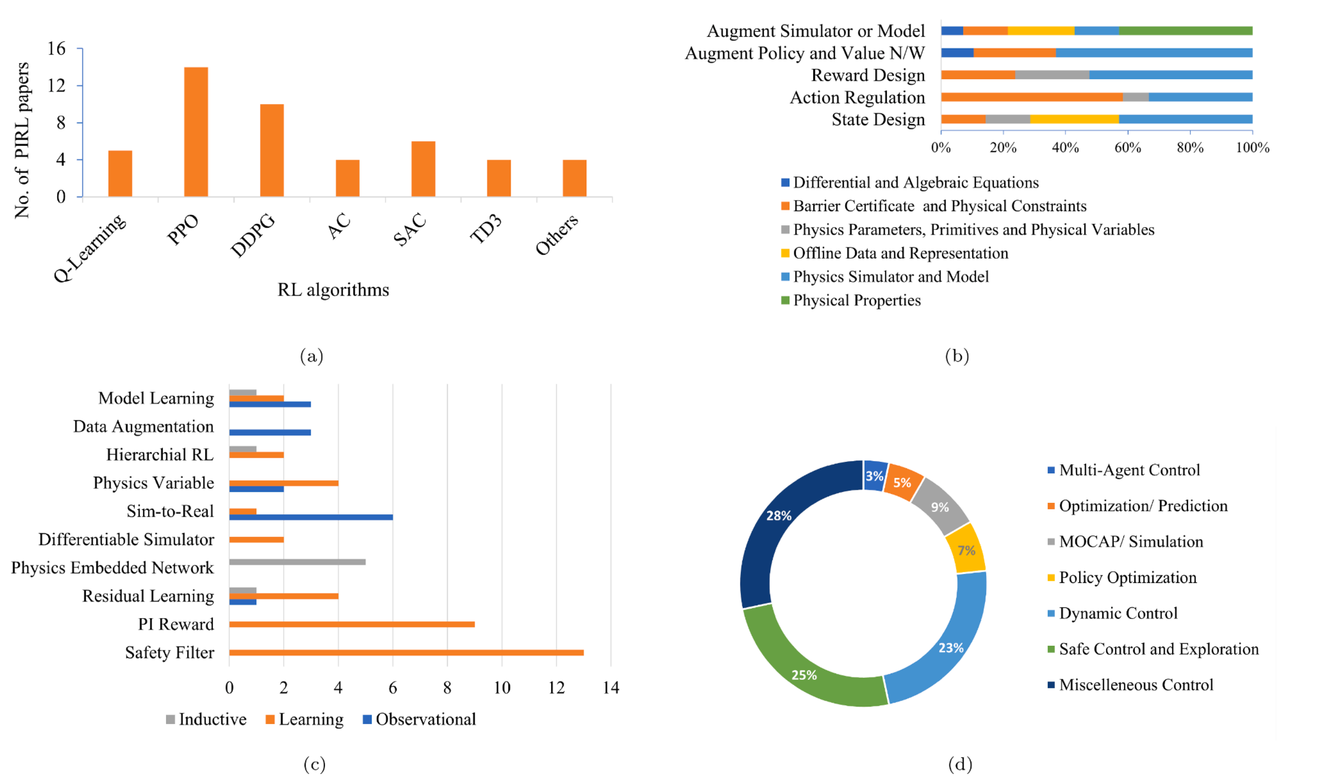
Types of physics priors used: In Fig. 15 (b), we can see that physics information takes the form of physics simulator, system models, barrier certificates and physical constraints, in a majority of works. PI types "Barrier certificate constraints and physical constraint" and "Physics simulator and models" dominates in more that 60% of works in "Action regulation" and "Augment policy and value N/W" PIRL methods.

Learning architecture and bias: In Fig. 15 (c) we visualize the relationship between PIRL learning architectures (sec: 3.4.2) and the three biases through which physics is typically incorporated in PIML approaches. In architectures "PI reward" and "safety filter", physics is incorporated strictly through "learning bias", signifying the heavy use of constraints, regularizers and specialized loss functions. While "Physics embedded network" incorporates physics information through "inductive bias", i.e., through imposition of hard constraints through use specialized and custom physics embodied networks.

Application domains: In Fig. 15 (d) almost 85% of the application problem dealt with PIRL approaches relates to controller or policy design. "Miscellaneous control" includes optimal policy/ controller learning approaches for different application sectors like energy management (Li et al., 2023; She et al., 2023) and data-center cooling (Wang et al., 2023), and accounts to majority of applications. "Safe control and exploration", includes those works concerning with safety critical systems, ensuring safe exploration and policy learning, accounts for 25%. "Dynamic control", includes control of dynamic systems, including robot systems and amounts to about 23% of all works surveyed. Other specific applications include optimization/ prediction (Cho et al., 2019; Dang & Ishii, 2022), motion capture/simulation (Chentanez et al., 2018; Wang, 2022) and improvement of general policy optimization approaches (Golemo et al., 2018; Lowrey et al., 2018) through physics incorporation.

4.3.2. Impact on learning efficiency

Different physics priors demonstrate varying impacts on learning efficiency. Barrier Certificates and Control Constraints (BPC) (Cheng et al., 2019; Zhao et al., 2023) significantly reduce exploration requirements by constraining action spaces to physically safe regions, showing



(a) (b) (c) (d)

Fig. 15. Statistical analysis of PIRL literature. (a) Statistic of type of RL algorithms used, (b) Statistic of PI Types used in each PIRL method, (c) Statistic of PIRL Learning Architectures and related Biases (d) Statistic of PIRL Applications in different domains.

Table 4
Summary of PIRL training/ evaluation benchmarks.

Simulator/ platform	Specific environment/ system name	Reference
OpenAI Gym	Pusher, Striker, ErgoReacher	Golemo et al. (2018)
OpenAI Gym	Mountain Car, Lunar Lander (*continuous*)	Jiang et al. (2022)
OpenAI Gym	Cart-Pole, Pendulum (simple and double)	Xie et al. (2016)
OpenAI Gym	Cart-pole	Cao et al. (2023c)
OpenAI Gym	Cart-pole and Quadruped robot	Cao et al. (2023b)
OpenAI Gym	CartPole, Pendulum	Liu and Wang (2021)
OpenAI Gym	Inverted Pendulum ($pendulum - v0$),	Cheng et al. (2019)
OpenAI Gym	Mountain car (*cont.*), Pendulum, Cart pole	Zhao et al. (2022)
OpenAI Gym	Simulated car following (He, Jin, & Orosz, 2018)	
MuJoCo	Ant, HalfCheetah, Humanoid, Walker2d	Mukherjee and Liu (2023)
	Humanoid standup, Swimmer, Hopper	
	Inverted and Inverted Double Pendulum ($v4$)	
MuJoCo	Cassie-MuJoCo-sim (robotics, Year Published/ Last Updated)	Duan et al. (2021); Siekmann et al. (2021)
	6 DoF Kinova Jaco (Ghosh, Singh, Rajeswaran, Kumar, & Levine, 2017)	Bahl et al. (2021, 2020)
MuJoCo	HalfCheetah, Ant,	Lee et al. (2020)
	CrippledHalfCheetah, and SlimHumanoid (Zhou et al., 2018)	
MuJoCo	Block stacking task (Janner et al., 2018)	Veerapaneni et al. (2020)
OpenAI Gym	CartPole, Pendulum	
OpenSim-RL (Kidziński et al., 2018)	L2M2019 environment	Korivand et al. (2023)
Safety gym (Yuan et al., 2021)	Point, car and Doggo goal	Yang et al. (2023)
-	Cart pole swing up, Ant	Xu et al. (2022)
-	Humanoid, Humanoid MTU	
-	Autonomous driving system	Chen et al. (2023b)
Deep control suite (Tassa et al., 2018)	Pendulum, Cartpole, Walker2d	Sanchez-Gonzalez et al. (2018)
	Acrobot, Swimmer, Cheetah	
-	JACO arm (real world)	
Deep control suite	Reacher, Pendulum, Cartpole,	Ramesh and Ravindran (2023)
	Cart-2-pole, Acrobot,	
	Cart-3-pole and Acro-3-bot	
-	Rabbit (Chevallereau et al., 2003)	Choi et al. (2020)
MARL env. (Lowe et al., 2017)	Multi-agent particle env.	Cai et al. (2021)
ADROIT (Rajeswaran et al., 2017)	Shadow dexterous hand	(Garcia-Hernando et al., 2020)
-	First-Person Hand Action Benchmark (Garcia-Hernando, Yuan, Baek, & Kim, 2018)	
MuJoCo	Door opening, in-hand manipulation,	
	tool use and object relocation	
SUMO (Lopez et al., 2018),	-	(Han et al., 2022)
METANET (Kotsialos, Papageorgiou, Diakaki, Pavlis, & Middelham, 2002)		
SUMO	-	(Wang, 2022)
CARLA (Dosovitskiy, Ros, Codevilla, Lopez, & Koltun, 2017)	-	(Udatha et al., 2022)
Gazebo (Koenig & Howard, 2004)	Quadrotor ($IF750A$)	(Huang et al., 2023)
IEEE Distribution	IEEE 33-bus and 141-bus distr. N/W	(Chen et al., 2022)
system benchmarks	IEEE 33-node system	(Cao et al., 2023a; Chen et al., 2022)
	IEEE 9-bus standard system	(Gao et al., 2022)
-	Custom (COMSOL based)	(Alam et al., 2021)
-	Custom (DFT based)	(Cho et al., 2019)
-	Custom (based on (Vrettos, Kara, MacDonald, Andersson, & Callaway, 2016))	(Gokhale et al., 2022)
-	Custom (based on (Kesting, Treiber, Schönhof, Kranke, & Helbing, 2007))	(Jurj et al., 2021)
-	Custom	(Dang & Ishii, 2022; Li et al., 2023; Martin & Schaub, 2022)
-	Custom	(Park et al., 2023; Shi et al., 2023; Yin et al., 2023)
-	Custom	(Yousif et al., 2023; Yu et al., 2023; Zhao et al., 2023)
-	Custom	(Cohen & Belta, 2023; Li & Belta, 2019; Lutter et al., 2021)
-	Custom	(Chen et al., 2023b; Wang et al., 2023)
Open AI Gym	Custom (Reactor geometries)	Radaideh et al. (2021)
MATLAB-Simulink	Custom	(She et al., 2023; Zhang et al., 2022)
-	Custom (Zimmermann, Poranne, Bern, & Coros, 2019)	(Mora et al., 2021)
MATLAB	Cruise control	(Li et al., 2021)
Pygame	Custom	(Zhao & Liu, 2021)
-	Custom (Unicycle, Car-following)	(Emam, Notomista, Glotfelter, Kira, & Egerstedt, 2022)
-	Brushbot, Quadrotor (sim)	(Ohnishi et al., 2019)
	Phantom manipulation platform	(Lowrey et al., 2018)
Pybullet	2 finger gripper	
	gym-pybullet-drones(Panerati et al., 2021)	(Du et al., 2023)
Pybullet	Franka Panda, Flexiv Rizon	(Lv et al., 2022)
NimblePhysics(Werling, Omens, Lee, Exarchos, & Liu, 2021),		
-	Custom MOCAP	(Chentanez et al., 2018; Luo et al., 2020; Peng et al., 2018)

40-50% sample reduction while ensuring stability. Differential and Algebraic Equations (DAE) approaches (Gao et al., 2022; Mukherjee & Liu, 2023) exhibit the strongest sample efficiency, with PINN-HJB implementations (Zhang et al., 2022) reducing required samples by up to 70% compared to standard RL. Physics Simulators (PS) (Golemo et al., 2018; Lowrey et al., 2018) show moderate initial efficiency improvements but excel at reducing real-world samples (>85%) during deployment. Physics Parameters and Variables (PPV) (Jurj et al., 2021; Shi et al., 2023) provide 20-30% faster convergence alongside substantial robustness improvements. Most notably, approaches combining multiple physics incorporation methods (Cao et al., 2023b; Li & Belta, 2019) demonstrate superior efficiency gains across all metrics, with hybrid approaches like (Cao et al., 2023b) achieving both 65% sample reduction and enhanced generalization. This suggests that physics prior selection should be guided by specific efficiency objectives rather than adopting one-size-fits-all approaches.

4.3.3. Trends in PIRL research

Research Trends (2018-2023) Several clear trends emerge in this comprehensive survey of the PIRL literature. The field shows exponential growth since 2018 (Fig. 1), with a significant shift toward safety-critical applications utilizing control barrier functions and certificates (Cheng et al., 2019; Li et al., 2021; Zhao et al., 2023). There's increasing adoption of PPO as the preferred algorithm framework Fig. 15a and growing interest in dynamic control problems (Bahl et al., 2020; Ramesh & Ravindran, 2023). A portion of surveyed works show convergence toward combining multiple physics incorporation methods simultaneously (Cao et al., 2023b; Li & Belta, 2019) rather than focusing on a single approach. This multifaceted integration aligns with broader trends in physics-informed machine learning that emphasize physically plausible solutions through varied incorporation strategies.

Emerging Trends (2024) Since the drafting of this survey, Physics-Informed Reinforcement Learning (PIRL) has continued to evolve rapidly across diverse application domains. Recent research (2024) has expanded PIRL's reach into several emerging fields with significant real-world impact. In autonomous systems and robotics, PIRL has demonstrated superior performance in control optimization tasks, with applications ranging from tracking moving objects (Faria, Capron, Secchi, & de Souza, 2024) to advanced locomotion systems like physics-based musculoskeletal models (Ogum, Schomaker, & Carloni, 2024) and swimming in turbulent environments (Koh, Pagnier, & Chertkov, 2025). Energy systems research has leveraged PIRL for robust voltage control in electrical distribution networks (Wei et al., 2023), implementing physics-guided multi-agent frameworks (Chen, Liu, Wang, & Kamwa, 2023a), and managing probabilistic wind power prediction (Chen, Liu, Wang, & Kamwa, 2024). In healthcare applications, PIRL has enhanced rehabilitation by optimizing arm movements through co-kriging adjustment in functional electrical stimulation systems (Liu, Wang, & Liu, 2024) and developing hand-object interaction controllers (Wannawas, Diaz-Pintado, Narayan, & Faisal, 2024) for human-robot interactions. Transportation safety has become another critical application area, with implementations including real-time optimal traffic routing under uncertainties (Ke, Zou, Liu, & Qian, 2025) and safe multi-agent collision avoidance (Feng et al., 2024). Advanced modeling techniques have progressed through hybrid planning models (Asri, Sigaud, & Thome, 2024) and physics-informed deep transfer reinforcement learning for complex systems (Zeng et al., 2024). The most promising emerging trend appears to be in safety guarantees, where physics-model-guided worst-case sampling (Cao, Mao, Sha, & Caccamo, 2024) and domain knowledge integration have allowed for enhanced efficiency while maintaining safety assurances in complex dynamic environments. Multi-agent collaborative systems for distributed energy management (Chen et al., 2024) represent another frontier where PIRL continues to demonstrate significant advantages over traditional approaches.

4.3.4. RL challenges addressed

In this section we will discuss and elaborate on how recent physics incorporation in RL algorithms have addressed certain open problems of the RL paradigm.

1) *Sample efficiency:* RL approaches need a huge number of agent-environment interaction and related data to work. One effective way of dealing with this problem is to use a surrogate for the real environment in the form of a simulator or learned model via data-driven approaches.

   PIRL approaches incorporate physics to augment simulators thus reducing the sim-to-real gap, thereby bringing down online evaluation cycles (Alam et al., 2021; Lv et al., 2022). Also physics incorporation during system identification or model learning phase in MBRL help reduce sample efficiency through learning a truer to real environment using lesser training samples (Sanchez-Gonzalez et al., 2018; Veerapaneni et al., 2020).

2) *Curse of dimensionality:* RL algorithms become less efficient both in training and performing on environment defined with high-dimensional and continuous state and action spaces, known as the 'curse of dimensionality'. Typically dimensionality reduction techniques are used to encode the large state or action vectors into low dimensional representations. The RL algorithm is then trained in this low dimensional setting.

   PIRL approaches extract underlying physics information from environment through learning physically relevant low dimensional representation from high dimensional observation or state space (Cao et al., 2023a; Gokhale et al., 2022). In Gokhale et al. (2022), a PINN is utilized to extract physically relevant information about the system's hidden state, which is then used to learn a Q-function for policy optimization.

3) *Safety exploration:* Safe reinforcement learning involves learning control policies that guarantee system performance and respect safety constraints during both exploration and policy deployment.

   In safety-critical applications using reinforcement learning, it's crucial to regulate agent exploration. Control Lyapunov function (CLF) (Choi et al., 2020; Li & Belta, 2019), barrier certificate/ barrier function (BF), control barrier function/ certificate (CBF/ CBC) (Cai et al., 2021; Cheng et al., 2019) are commonly used concepts. Barrier certificates define safe states, while control barrier functions ensure states stay in the safety set. These approaches are typically used for systems with partial or learnable dynamics model and generally a known set of safe states/ actions.

4) *Partial observability or imperfect measurement:* Partial observability is a setting where due to noise, missing information, or outside interference, an RL agent is unable to obtain the complete states needed to understand the environment.

   PIRL approaches modify or enhance the state representation to provide more useful information, in cases of missing or inadequate information. This may involve state fusion, which incorporates additional physics or geographical information from the environment (Jurj et al., 2021) or other agents (Shi et al., 2023).

5) *Under-defined reward function:* Defining the reward function is critical in creating MDPs and ensuring the effectiveness and efficiency of RL algorithms. However, since they are created by humans, there is a risk of them being under-defined and not guiding the RL algorithm effectively in policy optimization.

   PIRL approaches introduce physics information through effective reward design or augmentation of existing reward functions with bonuses or penalties (Dang & Ishii, 2022; Garcia-Hernando et al., 2020; Luo et al., 2020; Siekmann et al., 2021). For example, in a sim-to-real setting, Siekmann et al. (2021) proposed a framework for specifying rewards that combines probabilistic costs associated with primary forces and velocities. The framework creates a parametric reward function for common robotic gaits, in biped robots.

5. Open challenges and research directions

5.1. High dimensional spaces

A large number of real world tasks deals with high dimensional and continuous state and action spaces. One popular method to address this high dimensionality issue is to compress the state space (or action space) vectors into low dimensional vectors. A PI based approach may learn high quality environment representations using deep networks and extract physically relevant low dimensional features from them.

Learning a compressed, informative latent space from high-dimensional continuous states (or actions) remains a challenge, particularly in ensuring physical relevance. Future research can leverage Physics-Informed Autoencoders, which embed PDE constraints into loss functions (Yang, Rosca, Narasimhan, & Ramadge, 2022), and enforce physics-aware structures during training (Vatellis, 2024). Additionally, Latent Diffusion Models can be utilized to generate structured representations, which can then be constrained by physical laws (Shmakov et al., 2023).

5.2. Safety in complex and uncertain environments

In the realm of safe reinforcement learning, striking a balance between the complexity of the environment and ensuring safety is always a challenge. Current physics informed approaches uses different control theoretic concepts e.g. CBFs to ensure safe exploration and learning of the RL agent. But these approaches are limited by the approximated model of the system and the prior knowledge about safe state sets. There has been a lot of research for better system identification or model learning through physics incorporation. But most works do not generalize well to different tasks and environments.

To summarize, future research should focus on two key goals: 1) advancing model-agnostic safe exploration and control for RL agents in complex, uncertain environments, and 2) developing generalized methods for integrating physics into data-driven model learning. One promising direction is (Tayal, Singh, Kolathaya, & Bansal, 2025) which addresses both these objectives by formulating the co-optimization of safety and performance as a state-constrained optimal control problem, using Hamilton-Jacobi-Bellman (HJB) equations to efficiently approximate value functions. Additionally, it introduces conformal prediction-based verification to ensure generalization across complex, high-dimensional environments, effectively balancing safety and performance in autonomous systems.

5.3. Choice of physics prior

The choice of the physics prior is critical for the PIRL algorithm, yet it is challenging as it requires extensive study and can vary significantly across different systems, even within the same domain. This variability can hinder the efficiency of PIRL. To address this, a more comprehensive framework that incorporates physics information for managing novel physical tasks is preferred, rather than tackling tasks individually. An exemplar direction is the use of Physics-Guided Foundation Models (PGFMs) (Farhadloo et al., 2025), which integrate broad-domain physical knowledge to enhance model robustness, generalization, and prediction reliability across diverse systems.

5.4. Evaluation and benchmarking platform

Currently, PIRL doesn't have comprehensive benchmarking and evaluation environments to test and compare new physics approaches before induction. This limitation makes it challenging to assess the quality and uniqueness of new works.

Additionally, most PIRL works rely on customized environments related to a particular domain, making it difficult to compare PIRL algorithms fairly. Moreover, PIRL application cases are diverse, and the physics information chosen is specific to a domain, requiring extensive study and domain expertise to understand and compare such works.

5.5. Scalability for large-scale problems

A significant challenge for PIRL approaches is scaling to large, complex systems. Current methods face computational bottlenecks when physics-informed components like barrier certificates (Cheng et al., 2019; Zhao et al., 2023) and differentiable simulators (Xu et al., 2022) are applied to high-dimensional state spaces. As system complexity increases, maintaining accurate physics models becomes more difficult, with approximation errors potentially cascading through the learning process (Ramesh & Ravindran, 2023). Several promising directions has been dicussed that address these challenges: hierarchical decomposition approaches (Bahl et al., 2021) that break down complex problems into manageable subproblems; reduced-order modeling techniques (Veerapaneni et al., 2020) that abstract unnecessary details while preserving essential dynamics; physics-guided representation learning (Cao et al., 2023a; Gokhale et al., 2022) for dimensionality reduction; and distributed PIRL architectures (Shi et al., 2023) for multi-agent systems. Recent advancements in adaptive physics fidelity frameworks such as the Hyper-Low-Rank PINN method (Torres, Schiefer, & Niepert, 2025) have shown promising results in reducing computational costs while preserving accuracy through SVD-based weight decomposition. Approaches like the Enhanced Hybrid Adaptive PINN (Luo, Liao, Guan, & Liu, 2025) with dynamic collocation point allocation and the Difficulty-Aware Task Sampler (Toloubidokhti et al., 2023) that addresses performance disparities through meta-learning represent crucial research directions to enable PIRL's transition from controlled experimental settings to complex real-world applications.

6. Conclusions

This paper presents a state-of-the-art reinforcement learning paradigm, known as physics-informed reinforcement learning (PIRL). By leveraging both data-driven techniques and knowledge of underlying physical principles, PIRL is capable of improving the effectiveness, sample efficiency and accelerated training of RL algorithms/ approaches, for complex problem-solving and real-world deployment. We have created two taxonomies that categorize conventional PIRL methods based on physics prior/information type and physics prior induction (RL methods), providing a framework for understanding this approach. To help readers comprehend the physics involved in solving RL tasks, we have included various explanatory images from recent papers and summarized their characteristics in Tables 2 and 3. Additionally, we have provided a benchmark-summary Table 4 detailing the training and evaluation benchmarks used for PIRL evaluation. Our objective is to simplify the complex concepts of existing PIRL approaches, making them more accessible for use in various domains. Finally, we discuss the limitations and unanswered questions of current PIRL work, encouraging further research in this area.

Data availability

No data was used for the research described in the article.

CRediT authorship contribution statement

Chayan Banerjee: Conceptualization, Investigation, Visualization, Writing – original draft, Writing – review & editing; **Kien Nguyen:** Conceptualization, Supervision, Writing – review & editing; **Clinton Fookes:** Supervision; **Maziar Raissi:** Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr.

Kien Nguyen reports financial support was provided by Queensland Government Advance Queensland.

Acknowledgment

This research was partly supported by the Advance Queensland Industry Research Fellowship AQIRF024-2021RD4.

References

Alam, M. F., Shtein, M., Barton, K., & Hoelzle, D. J. (2021). A physics-guided reinforcement learning framework for an autonomous manufacturing system with expensive data. In *2021 American control conference (ACC)* (pp. 484–490).

Ames, A. D., Coogan, S., Egerstedt, M., Notomista, G., Sreenath, K., & Tabuada, P. (2019). Control barrier functions: Theory and applications. In *2019 18th European control conference (ECC)* (pp. 3420–3431).

Asri, Z. E., Sigaud, O., & Thome, N. (2024). Physics-informed model and hybrid planning for efficient dyna-style reinforcement learning. *arXiv preprint arXiv:2407.02217*.

Bahl, S., Gupta, A., & Pathak, D. (2021). Hierarchical neural dynamic policies. *arXiv preprint arXiv:2107.05627*.

Bahl, S., Mukadam, M., Gupta, A., & Pathak, D. (2020). Neural dynamic policies for end-to-end sensorimotor learning. *Advances in Neural Information Processing Systems*, *33*, 5058–5069.

Banerjee, C., Chen, Z., & Noman, N. (2022a). Boosting exploration in actor-critic algorithms by incentivizing plausible novel states. *arXiv preprint arXiv:2210.00211*.

Banerjee, C., Chen, Z., & Noman, N. (2022b). Improved soft actor-critic: Mixing prioritized off-policy samples with on-policy experiences. *IEEE Transactions on Neural Networks and Learning Systems*.

Banerjee, C., Chen, Z., Noman, N., & Zamani, M. (2022c). Optimal actor-critic policy with optimized training datasets. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *6*(6), 1324–1334.

Banerjee, C., Nguyen, K., Fookes, C., & Karniadakis, G. (2023). Physics-informed computer vision: A review and perspectives.

Barth-Maron, G., Hoffman, M. W., Budden, D., Dabney, W., Horgan, D., Tb, D., Muldal, A., Heess, N., & Lillicrap, T. (2018). Distributed distributional deterministic policy gradients. *arXiv preprint arXiv:1804.08617*.

Booth, S., Knox, W. B., Shah, J., Niekum, S., Stone, P., & Allievi, A. (2023). The perils of trial-and-error reward design: Misdesign through overfitting and invalid task specifications. In *Aaai conference on artificial intelligence* (pp. 5920–5929). (*vol. 37*).

Cai, Z., Cao, H., Lu, W., Zhang, L., & Xiong, H. (2021). Safe multi-agent reinforcement learning through decentralized multiple control barrier functions. *arXiv preprint arXiv:2103.12553*.

Cao, D., Zhao, J., Hu, J., Pei, Y., Huang, Q., Chen, Z., & Hu, W. (2023a). Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control. *IEEE Transactions on Smart Grid*.

Cao, H., Mao, Y., Sha, L., & Caccamo, M. (2023b). Physical deep reinforcement learning: Safety and unknown unknowns. *arXiv preprint arXiv:2305.16614*.

Cao, H., Mao, Y., Sha, L., & Caccamo, M. (2023c). Physical deep reinforcement learning towards safety guarantee. *arXiv preprint arXiv:2303.16860*.

Cao, H., Mao, Y., Sha, L., & Caccamo, M. (2024). Physics-model-guided worst-case sampling for safe reinforcement learning. *arXiv preprint arXiv:2412.13224*.

Chen, C., Lewis, F. L., Xie, K., Xie, S., & Liu, Y. (2020). Off-policy learning for adaptive optimal output synchronization of heterogeneous multi-agent systems. *Automatica*, *119*, 109081.

Chen, J., Yuan, B., & Tomizuka, M. (2019). Model-free deep reinforcement learning for urban autonomous driving. In *2019 IEEE intelligent transportation systems conference (ITSC)* (pp. 2765–2771).

Chen, P., Liu, S., Wang, X., & Kamwa, I. (2022). Physics-shielded multi-agent deep reinforcement learning for safe active voltage control with photovoltaic/battery energy storage systems. *IEEE Transactions on Smart Grid*.

Chen, P., Liu, S., Wang, X., & Kamwa, I. (2023a). Physics-guided multi-agent deep reinforcement learning for robust active voltage control in electrical distribution systems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *71*(2), 922–933.

Chen, P., Liu, S., Wang, X., & Kamwa, I. (2024). Physics-guided multi-agent adversarial reinforcement learning for robust active voltage control with peer-to-peer (p2p) energy trading. *IEEE Transactions on Power Systems*.

Chen, X., Liu, S., & Di, X. (2023b). A hybrid framework of reinforcement learning and physics-informed deep learning for spatiotemporal mean field games. In *2023 International conference on autonomous agents and multiagent systems* (pp. 1079–1087).

Cheng, R., Orosz, G., Murray, R. M., & Burdick, J. W. (2019). End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *AAAI conference on artificial intelligence* (pp. 3387–3395). (*vol. 33*).

Chentanez, N., Müller, M., Macklin, M., Makoviychuk, V., & Jeschke, S. (2018). Physics-based motion capture imitation with deep reinforcement learning. In *11th ACM SIGGRAPH conference on motion, interaction and games* (pp. 1–10).

Chevallereau, C., Abba, G., Plestan, F., Westervelt, E., de Wit, C. C., Grizzle, J. et al. (2003). Rabbit: A testbed for advanced control theory. *IEEE Control Systems Magazine*, *23*(5), 57–79.

Cho, Y., Kim, S., Li, P. P., Surh, M. P., Han, T. Y.-J., & Choo, J. (2019). Physics-guided reinforcement learning for 3d molecular structures. In *Workshop at the 33rd conference on neural information processing systems (neurIPS)*.

Choi, J., Castaneda, F., Tomlin, C. J., & Sreenath, K. (2020). Reinforcement learning for safety-critical control under model uncertainty, using control lyapunov functions and control barrier functions. *arXiv preprint arXiv:2004.07584*.

Cohen, M. H., & Belta, C. (2023). Safe exploration in model-based reinforcement learning using control barrier functions. *Automatica*, *147*, 110684.

Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., & Ho, S. (2020). Lagrangian neural networks. *arXiv preprint arXiv:2003.04630*.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. *arXiv preprint arXiv:2201.05624*.

Cutler, M., Walsh, T. J., & How, J. P. (2015). Real-world reinforcement learning via multifidelity simulators. *IEEE Transactions on Robotics*, *31*(3), 655–671.

Dang, Z., & Ishii, M. (2022). Towards stochastic modeling for two-phase flow interfacial area predictions: A physics-informed reinforcement learning approach. *International Journal of Heat and Mass Transfer*, *192*, 122919.

Dawson, C., Qin, Z., Gao, S., & Fan, C. (2022). Safe nonlinear control using robust neural lyapunov-barrier functions. In *Conference on robot learning* (pp. 1724–1735). PMLR.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). Carla: An open urban driving simulator. In *Conference on robot learning* (pp. 1–16). PMLR.

Du, D., Han, S., Qi, N., Ammar, H. B., Wang, J., & Pan, W. (2023). Reinforcement learning for safe robot control using control lyapunov barrier functions. *arXiv preprint arXiv:2305.09793*.

Duan, H., Dao, J., Green, K., Apgar, T., Fern, A., & Hurst, J. (2021). Learning task space actions for bipedal locomotion. In *2021 IEEE International conference on robotics and automation (ICRA)* (pp. 1276–1282).

Dulac-Arnold, G., Evans, R., van Hasselt, H., Sunehag, P., Lillicrap, T., Hunt, J., Mann, T., Weber, T., Degris, T., & Coppin, B. (2015). Deep reinforcement learning in large discrete action spaces. *arXiv preprint arXiv:1512.07679*.

Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Gowal, S., & Hester, T. (2021). Challenges of real-world reinforcement learning: Definitions, benchmarks and analysis. *Machine Learning*, *110*(9), 2419–2468.

EEßerer, J., Bach, N., Jestel, C., Urbann, O., & Kerner, S. (2022). Guided reinforcement learning: A review and evaluation for efficient and effective real-world robotics. *IEEE Robotics & Automation Magazine*.

Emam, Y., Notomista, G., Glotfelter, P., Kira, Z., & Egerstedt, M. (2022). Safe reinforcement learning using robust control barrier functions. *IEEE Robotics and Automation Letters*, (99), 1–8.

Erichson, N. B., Muehlebach, M., & Mahoney, M. W. (2019). Physics-informed autoencoders for lyapunov-stable fluid flow prediction. *arXiv preprint arXiv:1905.10866*.

Farhadloo, M., Sharma, A., Yang, M., Jayaprakash, B., Northrop, W., & Shekhar, S. (2025). Towards physics-guided foundation models. *arXiv preprint arXiv:2502.15013*.

Faria, R. R., Capron, B., Secchi, A. R., & de Souza, M. B. Jr (2024). A data-driven tracking control framework using physics-informed neural networks and deep reinforcement learning for dynamical systems. *Engineering Applications of Artificial Intelligence*, *127*, 107256.

Feng, P., Shi, R., Wang, S., Liang, J., Yu, X., Li, S., & Wu, W. (2024). Safe and efficient multi-agent collision avoidance with physics-informed reinforcement learning. *IEEE Robotics and Automation Letters*.

Fujimoto, S., Hoof, H., & Meger, D. (2018). Addressing function approximation error in actor-critic methods. In *International conference on machine learning* (pp. 1587–1596). PMLR.

Gao, B., & Pavel, L. (2021). On passivity, reinforcement learning, and higher order learning in multiagent finite games. *IEEE Transactions on Automatic Control*, *66*(1), 121–136.

Gao, J., Chen, S., Li, X., & Zhang, J. (2022). Transient voltage control based on physics-informed reinforcement learning. *IEEE Journal of Radio Frequency Identification*, *6*, 905–910.

Garcia, J., & Fernández, F. (2012). Safe exploration of state and action spaces in reinforcement learning. *Journal of Artificial Intelligence Research*, *45*, 515–564.

Garcia-Hernando, G., Johns, E., & Kim, T.-K. (2020). Physics-based dexterous manipulations with estimated hand poses and residual reinforcement learning. In *2020 IEEE/RSJ International conference on intelligent robots and systems (IROS)* (pp. 9561–9568).

Garcia-Hernando, G., Yuan, S., Baek, S., & Kim, T.-K. (2018). First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In *IEEE conference on computer vision and pattern recognition* (pp. 409–419).

Ghosh, D., Singh, A., Rajeswaran, A., Kumar, V., & Levine, S. (2017). Divide-and-conquer reinforcement learning. *arXiv preprint arXiv:1711.09874*.

Gokhale, G., Claessens, B., & Develder, C. (2022). Physq: A physics informed reinforcement learning framework for building control. *arXiv preprint arXiv:2211.11830*.

Golemo, F., Taiga, A. A., Courville, A., & Oudeyer, P.-Y. (2018). Sim-to-real transfer with neural-augmented robot simulation. In *Conference on robot learning* (pp. 817–828). PMLR.

Greydanus, S., Dzamba, M., & Yosinski, J. (2019). Hamiltonian neural networks. *Advances in Neural Information Processing Systems*, *32*.

Gu, S., Yang, L., Du, Y., Chen, G., Walter, F., Wang, J., Yang, Y., & Knoll, A. (2022). A review of safe reinforcement learning: Methods, theory and applications. *arXiv preprint arXiv:2205.10330*.

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning* (pp. 1861–1870). PMLR.

Hafner, D., Lillicrap, T., Ba, J., & Norouzi, M. (2019). Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*.

Han, Y., Wang, M., Li, L., Roncoli, C., Gao, J., & Liu, P. (2022). A physics-informed reinforcement learning-based strategy for local and coordinated ramp metering. *Transportation Research Part C: Emerging Technologies*, *137*, 103584.

Hao, Z., Liu, S., Zhang, Y., Ying, C., Feng, Y., Su, H., & Zhu, J. (2022). Physics-informed machine learning: A survey on problems, methods and applications. *arXiv preprint arXiv:2211.08064*.

He, C. R., Jin, I. G., & Orosz, G. (2018). Data-based fuel-economy optimization of connected automated trucks in traffic. In *2018 annual American control conference (ACC)* (pp. 5576–5581).

Huang, J., Zeng, W., Xiong, H., Noack, B. R., Hu, G., Liu, S., Xu, Y., & Cao, H. (2023). Symmetry-informed reinforcement learning and its application to the attitude control of quadrotors. *IEEE Transactions on Artificial Intelligence*.

Janner, M., Levine, S., Freeman, W. T., Tenenbaum, J. B., Finn, C., & Wu, J. (2018). Reasoning about physical interactions with object-oriented prediction and planning. *arXiv preprint arXiv:1812.10972*.

Jiang, J., Fu, M., & Chen, Z. (2022). Physics informed intrinsic rewards in reinforcement learning. In *2022 Australian & New Zealand control conference (ANZCC)* (pp. 74–69).

Jiang, Y., Fan, J., Gao, W., Chai, T., & Lewis, F. L. (2020). Cooperative adaptive optimal output regulation of nonlinear discrete-time multi-agent systems. *Automatica*, *121*, 109149.

Johannink, T., Bahl, S., Nair, A., Luo, J., Kumar, A., Loskyll, M., Ojea, J. A., Solowjow, E., & Levine, S. (2019). Residual reinforcement learning for robot control. In *2019 International conference on robotics and automation (ICRA)* (pp. 6023–6029).

Jurj, S. L., Grundt, D., Werner, T., Borchers, P., Rothemann, K., & Möhlmann, E. (2021). Increasing the safety of adaptive cruise control using physics-guided reinforcement learning. *Energies*, *14*(22), 7572.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440.

Kashefi, A., Rempe, D., & Guibas, L. J. (2021). A point-cloud deep learning framework for prediction of fluid flow fields on irregular geometries. *Physics of Fluids*, *33*(2), 027104.

Kashinath, K., Mustafa, M., Albert, A., Wu, J. L., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., Singh, A. et al. (2021). Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A*, *379*, 1–36.

Ke, Z., Zou, Q., Liu, J., & Qian, S. (2025). Real-time system optimal traffic routing under uncertainties-can physics models boost reinforcement learning? *Transportation Research Part C: Emerging Technologies*, *173*, 105040.

Kesting, A., Treiber, M., Schönhof, M., Kranke, F., & Helbing, D. (2007). Jam-avoiding adaptive cruise control (ACC) and its impact on traffic dynamics. In *Traffic and granular flow'05* (pp. 633–643). Springer.

Kidziński, Ł., Mohanty, S. P., Ong, C. F., Hicks, J. L., Carroll, S. F., Levine, S., Salathé, M., & Delp, S. L. (2018). Learning to run challenge: Synthesizing physiologically accurate motion using deep reinforcement learning. In *The NIPS'17 competition: Building intelligent systems* (pp. 101–120). Springer.

Knox, W. B., Allievi, A., Banzhaf, H., Schmitt, F., & Stone, P. (2023). Reward (mis) design for autonomous driving. *Artificial Intelligence*, *316*, 103829.

Koenig, N., & Howard, A. (2004). Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International conference on intelligent robots and systems (IROS)(IEEE cat. no. 04CH37566)* (pp. 2149–2154). (*vol. 3*).

Koh, C., Pagnier, L., & Chertkov, M. (2025). Physics-guided actor-critic reinforcement learning for swimming in turbulence. *Physical Review Research*, *7*(1), 013121.

Konda, V., & Tsitsiklis, J. (1999). Actor-critic algorithms. *Advances in Neural Information Processing Systems*, *12*.

Korivand, S., Jalili, N., & Gong, J. (2023). Inertia-constrained reinforcement learning to enhance human motor control modeling. *Sensors*, *23*(5), 2698.

Kotsialos, A., Papageorgiou, M., Diakaki, C., Pavlis, Y., & Middelham, F. (2002). Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET. *IEEE Transactions on Intelligent Transportation Systems*, *3*(4), 282–292.

Lee, K., Seo, Y., Lee, S., Lee, H., & Shin, J. (2020). Context-aware dynamics model for generalization in model-based reinforcement learning. In *International conference on machine learning* (pp. 5757–5766). PMLR.

Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, *17*(1), 1334–1373.

Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*.

Li, X., & Belta, C. (2019). Temporal logic guided safe reinforcement learning using control barrier functions. *arXiv preprint arXiv:1903.09885*.

Li, Y., He, S., Li, Y., Shi, Y., & Zeng, Z. (2023). Federated multiagent deep reinforcement learning approach via physics-informed reward for multimicrogrid energy management. *IEEE Transactions on Neural Networks and Learning Systems*.

Li, Y., Li, N., Tseng, H. E., Girard, A., Filev, D., & Kolmanovsky, I. (2021). Safe reinforcement learning using robust action governor. In *Learning for dynamics and control* (pp. 1093–1104). PMLR.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.

Liu, X.-Y., & Wang, J.-X. (2021). Physics-informed dyna-style model-based deep reinforcement learning for dynamic control. *Royal Society A*, *477*(2255), 20210618.

Liu, Y., Wang, J., & Liu, L. (2024). Physics-informed reinforcement learning for probabilistic wind power forecasting under extreme events. *Applied Energy*, *376*, 124068.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wießner, E. (2018). Microscopic traffic simulation using sumo. In *2018 21st International conference on intelligent transportation systems (ITSC)* (pp. 2575–2582).

Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems*, *30*.

Lowrey, K., Kolev, S., Dao, J., Rajeswaran, A., & Todorov, E. (2018). Reinforcement learning for non-prehensile manipulation: Transfer from simulation to physical system. In *2018 IEEE International conference on simulation, modeling, and programming for autonomous robots (SIMPAR)* (pp. 35–42).

Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via deepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, *3*(3), 218–229.

Luo, K., Liao, S., Guan, Z., & Liu, B. (2025). An enhanced hybrid adaptive physics-informed neural network for forward and inverse PDE problems. *Applied Intelligence*, *55*(3), 1–20.

Luo, Z., Hachiuma, R., Yuan, Y., Iwase, S., & Kitani, K. M. (2020). Kinematics-guided reinforcement learning for object-aware 3d ego-pose estimation. *arXiv preprint arXiv:2011.04837*.

Lutter, M., Silberbauer, J., Watson, J., & Peters, J. (2021). Differentiable physics models for real-world offline model-based reinforcement learning. In *2021 IEEE International conference on robotics and automation (ICRA)* (pp. 4163–4170).

Lv, J., Feng, Y., Zhang, C., Zhao, S., Shao, L., & Lu, C. (2022). Sam-rl: Sensing-aware model-based reinforcement learning via differentiable physics-based simulation and rendering. *arXiv preprint arXiv:2210.15185*.

Ma, H., Chen, J., Eben, S., Lin, Z., Guan, Y., Ren, Y., & Zheng, S. (2021). Model-based constrained reinforcement learning using generalized control barrier function. In *2021 IEEE/RSJ International conference on intelligent robots and systems (IROS)* (pp. 4552–4559).

Margolis, G. B., Chen, T., Paigwar, K., Fu, X., Kim, D., Kim, S., & Agrawal, P. (2021). Learning to jump from pixels. *arXiv preprint arXiv:2110.15344*.

Martin, J., & Schaub, H. (2022). Reinforcement learning and orbit-discovery enhanced by small-body physics-informed neural network gravity models. In *AIAA scitech 2022 forum* (p. 2272).

Meng, C., Seo, S., Cao, D., Griesemer, S., & Liu, Y. (2022a). When physics meets machine learning: A survey of physics-informed machine learning. *arXiv preprint arXiv:2203.16797*.

Meng, X., Yang, L., Mao, Z., del Águila Ferrandis, J., & Karniadakis, G. E. (2022b). Learning functional priors and posteriors from data and physics. *Journal of Computational Physics*, *457*, 111073.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Mora, M. A. Z., Peychev, M., Ha, S., Vechev, M., & Coros, S. (2021). Pods: Policy optimization via differentiable simulation. In *International conference on machine learning* (pp. 7805–7817). PMLR.

Mukherjee, A., & Liu, J. (2023). Bridging physics-informed neural networks with reinforcement learning: Hamilton-jacobi-bellman proximal policy optimization (HJBPPO). *arXiv preprint arXiv:2302.00237*.

Neunert, M., Abdolmaleki, A., Wulfmeier, M., Lampe, T., Springenberg, T., Hafner, R., Romano, F., Buchli, J., Heess, N., & Riedmiller, M. (2020). Continuous-discrete reinforcement learning for hybrid control in robotics. *Conference on Robot Learning*, (pp. 735–751).

Ogum, B. N., Schomaker, L. R. B., & Carloni, R. (2024). Learning to walk with deep reinforcement learning: forward dynamic simulation of a physics-based musculoskeletal model of an osseointegrated transfemoral amputee. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *32*, 431–441.

Ohnishi, M., Wang, L., Notomista, G., & Egerstedt, M. (2019). Barrier-certified adaptive reinforcement learning with applications to brushbot navigation. *IEEE Transactions on Robotics*, *35*(5), 1186–1205.

Osiński, B., Jakubowski, A., Zięcina, P., Miłoś, P., Galias, C., Homoceanu, S., & Michalewski, H. (2020). Simulation-based reinforcement learning for real-world autonomous driving. In *2020 IEEE International conference on robotics and automation (ICRA)* (pp. 6411–6418).

Panerati, J., Zheng, H., Zhou, S., Xu, J., Prorok, A., & Schoellig, A. P. (2021). Learning to fly-a gym environment with pybullet physics for reinforcement learning of multi-agent quadcopter control. In *2021 IEEE/RSJ International conference on intelligent robots and systems (IROS)* (pp. 7512–7519).

Park, C., Kim, S., Jung, A. W., Park, J., Seo, D., Kim, Y., Park, C., Park, C. Y., & Jang, M. S. (2023). Physics-informed reinforcement learning for sample-efficient optimization of freeform nanophotonic devices. *arXiv preprint arXiv:2306.04108*.

Peng, X. B., Abbeel, P., Levine, S., & Van de Panne, M. (2018). Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, *37*(4), 1–14.

Peng, X. B., Berseth, G., Yin, K., & Van De Panne, M. (2017). DeepLoco: Dynamic locomotion skills using hierarchical deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, *36*(4), 1–13.

Radaideh, M. I., Wolverton, I., Joseph, J., Tusar, J. J., Otgonbaatar, U., Roy, N., Forget, B., & Shirvan, K. (2021). Physics-informed reinforcement learning optimization of nuclear assembly design. *Nuclear Engineering and Design*, *372*, 110966.

Rajeswaran, A., Kumar, V., Gupta, A., Vezzani, G., Schulman, J., Todorov, E., & Levine, S. (2017). Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. *arXiv preprint arXiv:1709.10087*.

Ramesh, A., & Ravindran, B. (2023). Physics-informed model-based reinforcement learning. In *Learning for dynamics and control conference* (pp. 26–37). PMLR.

Riedmiller, M. (2005). Neural fitted q iteration–first experiences with a data efficient neural reinforcement learning method. In *16th European conference on machine learning* (pp. 317–328). Springer.

robotics, A. (Year Published/ Last Updated). Cassie-mujoco-sim. https://github.com/osu-drl/cassie-mujoco-sim.

Rodwell, C., & Tallapragada, P. (2023). Physics-informed reinforcement learning for motion control of a fish-like swimming robot. *Scientific Reports*, *13*(1), 1–17.

Romdlony, M. Z., & Jayawardhana, B. (2016). Stabilization with guaranteed safety using control lyapunov–barrier function. *Automatica*, *66*, 39–47.

Sanchez-Gonzalez, A., Heess, N., Springenberg, J. T., Merel, J., Riedmiller, M., Hadsell, R., & Battaglia, P. (2018). Graph networks as learnable physics engines for inference and control. In *International conference on machine learning* (pp. 4470–4479). PMLR.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Shah, V., Joshi, A., Ghosal, S., Pokuri, B., Sarkar, S., Ganapathysubramanian, B., & Hegde, C. (2019). Encoding invariances in deep generative models. *arXiv preprint arXiv:1906.01626*.

She, B., Li, F., Cui, H., Shuai, H., Oboreh-Snapps, O., Bo, R., Praisuwanna, N., Wang, J., & Tolbert, L. M. (2023). Inverter PQ control with trajectory tracking capability for microgrids based on physics-informed reinforcement learning. *IEEE Transactions on Smart Grid*.

Shi, H., Zhou, Y., Wu, K., Chen, S., Ran, B., & Nie, Q. (2023). Physics-informed deep reinforcement learning-based integrated two-dimensional car-following control strategy for connected automated vehicles. *Knowledge-Based Systems*, *269*, 110485.

Shmakov, A., Greif, K., Fenton, M., Ghosh, A., Baldi, P., & Whiteson, D. (2023). End-to-end latent variational diffusion models for inverse problems in high energy physics. *Advances in Neural Information Processing Systems*, *36*, 65102–65127.

Siekmann, J., Godse, Y., Fern, A., & Hurst, J. (2021). Sim-to-real learning of all common bipedal gaits via periodic reward composition. In *2021 IEEE International conference on robotics and automation (ICRA)* (pp. 7309–7315).

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., & Riedmiller, M. (2014). Deterministic policy gradient algorithms. In *International conference on machine learning* (pp. 387–395). Pmlr.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction MIT press Cambridge, MA, 22447.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., de Las, C. D., Budden, D., Abdolmaleki, A., Merel, J., Lefrancq, A. et al. (2018). Deepmind control suite. *arXiv preprint arXiv:1801.00690*.

Tayal, M., Singh, A., Kolathaya, S., & Bansal, S. (2025). A physics-informed machine learning framework for safe and optimal control of autonomous systems. *arXiv preprint arXiv:2502.11057*.

Tessler, C., Zahavy, T., Cohen, D., Mankowitz, D. J., & Mannor, S. (2019). Action assembly: Sparse imitation learning for text based games with combinatorial action spaces. *arXiv preprint arXiv:1905.09700*.

Toloubidokhti, M., Ye, Y., Missel, R., Jiang, X., Kumar, N., Shrestha, R., & Wang, L. (2023). Dats: Difficulty-aware task sampler for meta-learning physics-informed neural networks. In *The twelfth International conference on learning representations*.

Toromanoff, M., Wirbel, E., & Moutarde, F. (2020). End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *IEEE/CVF conference on computer vision and pattern recognition* (pp. 7153–7162).

Torres, E., Schiefer, J., & Niepert, M. (2025). Adaptive physics-informed neural networks: A survey. *arXiv preprint arXiv:2503.18181*.

Udatha, S., Lyu, Y., & Dolan, J. (2022). Safe reinforcement learning with probabilistic control barrier functions for ramp merging. *arXiv preprint arXiv:2212.00618*.

Vatellis, V. (2024). Advancing physics data analysis through machine learning and physics-informed neural networks. *arXiv preprint arXiv:2410.14760*.

Veerapaneni, R., Co-Reyes, J. D., Chang, M., Janner, M., Finn, C., Wu, J., Tenenbaum, J., & Levine, S. (2020). Entity abstraction in visual model-based reinforcement learning. In *Conference on robot learning* (pp. 1439–1456). PMLR.

Vrettos, E., Kara, E. C., MacDonald, J., Andersson, G., & Callaway, D. S. (2016). Experimental demonstration of frequency regulation by commercial buildings-part i: Modeling and hierarchical control design. *IEEE Transactions on Smart Grid*, *9*(4), 3213–3223.

Wang, R., Cao, Z., Zhou, X., Wen, Y., & Tan, R. (2023). Phyllis: Physics-informed lifelong reinforcement learning for data center cooling control. In *14th ACM International conference on future energy systems* (pp. 114–126).

Wang, X. (2022). Ensuring safety of learning-based motion planners using control barrier functions. *IEEE Robotics and Automation Letters*, *7*(2), 4773–4780.

Wannawas, N., Diaz-Pintado, C., Narayan, J., & Faisal, A. A. (2024). Controlling fes of arm movements using physics-informed reinforcement learning via co-kriging adjustment. In *2024 IEEE International conference on robotics and automation (ICRA)* (pp. 5555–5560). IEEE.

Wei, Z., Zhou, Z., Wang, P., Ren, J., Yin, Y., Pedersen, G. F., & Shen, M. (2023). Automated antenna design via domain knowledge-informed reinforcement learning and imitation learning. *IEEE Transactions on Antennas and Propagation*, *71*(7), 5549–5557.

Werling, K., Omens, D., Lee, J., Exarchos, I., & Liu, C. K. (2021). Fast and feature-complete differentiable physics for articulated rigid bodies with contact. *arXiv preprint arXiv:2103.16021*.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, *8*, 229–256.

Wu, J.-L., Kashinath, K., Albert, A., Chirila, D., Xiao, H. et al. (2020). Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems. *Journal of Computational Physics*, *406*, 109209.

Xie, C., Patil, S., Moldovan, T., Levine, S., & Abbeel, P. (2016). Model-based reinforcement learning with parametrized physical models and optimism-driven exploration. In *2016 IEEE International conference on robotics and automation (ICRA)* (pp. 504–511).

Xie, Z., Berseth, G., Clary, P., Hurst, J., & van de Panne, M. (2018). Feedback control for cassie with deep reinforcement learning. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (pp. 1241–1246).

Xu, J., Makoviychuk, V., Narang, Y., Ramos, F., Matusik, W., Garg, A., & Macklin, M. (2022). Accelerated policy learning with parallel differentiable simulation. *arXiv preprint arXiv:2204.07137*.

Yang, M., & Nachum, O. (2021). Representation matters: Offline pretraining for sequential decision making. In *International conference on machine learning* (pp. 11784–11794). PMLR.

Yang, T.-Y., Rosca, J., Narasimhan, K., & Ramadge, P. J. (2022). Learning physics constrained dynamics using autoencoders. *Advances in Neural Information Processing Systems*, *35*, 17157–17172.

Yang, Y., Jiang, Y., Liu, Y., Chen, J., & Li, S. E. (2023). Model-free safe reinforcement learning through neural barrier certificate. *IEEE Robotics and Automation Letters*.

Yang, Y., & Perdikaris, P. (2019). Conditional deep surrogate models for stochastic, high-dimensional, and multi-fidelity systems. *Computational Mechanics*, *64*(2), 417–434.

Yin, M., Li, T., Lei, H., Hu, Y., Rangan, S., & Zhu, Q. (2023). Generalizable wireless navigation through physics-informed reinforcement learning in wireless digital twin. *arXiv preprint arXiv:2306.06766*.

Yousif, M. Z., Zhang, M., Yang, Y., Zhou, H., Yu, L., & Lim, H. (2023). Physics-guided deep reinforcement learning for flow field denoising. *arXiv preprint arXiv:2302.09559*.

Yu, P., Zhang, H., & Song, Y. (2023). District cooling system control for providing regulation services based on safe reinforcement learning with barrier functions. *Applied Energy*, *347*, 121396.

Yuan, Z., Hall, A. W., Zhou, S., Brunke, L., Greeff, M., Panerati, J., & Schoellig, A. P. (2021). Safe-control-gym: A unified benchmark suite for safe learning-based control and reinforcement learning. *arXiv preprint arXiv:2109.06325*.

Zeng, Y., Xiao, Z., Liu, Q., Liang, G., Rodriguez, E., Zou, G., Zhang, X., & Pou, J. (2024). Physics-informed deep transfer reinforcement learning method for the input-series output-parallel dual active bridge-based auxiliary power modules in electrical aircraft. *IEEE Transactions on Transportation Electrification*.

Zhang, J., Mao, S., Yang, L., Ma, W., Li, S., & Gao, Z. (2024). Physics-informed deep learning for traffic state estimation based on the traffic flow model and computational graph method. *Information Fusion*, *101*, 101971.

Zhang, X., Peng, Y., Pan, W., Xu, X., & Xie, H. (2022). Barrier function-based safe reinforcement learning for formation control of mobile robots. In *2022 International conference on robotics and automation (ICRA)* (pp. 5532–5538).

Zhao, P., & Liu, Y. (2021). Physics informed deep reinforcement learning for aircraft conflict resolution. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 8288–8301.

Zhao, Q., Zhang, Y., & Li, X. (2022). Safe reinforcement learning for dynamical systems using barrier certificates. *Connection Science*, *34*(1), 2822–2844.

Zhao, T., Wang, J., & Yue, M. (2023). A barrier-certificated reinforcement learning approach for enhancing power system transient stability. *IEEE Transactions on Power Systems*.

Zheng, W., Yang, H. F., Cai, J., Wang, P., Jiang, X., Du, S. S., Wang, Y., & Wang, Z. (2023). Integrating the traffic science with representation learning for city-wide network congestion prediction. *Information Fusion*, *99*, 101837.

Zhou, W., Pinto, L., & Gupta, A. (2018). Environment probing interaction policies. In *International conference on learning representations*.

Zimmermann, S., Poranne, R., Bern, J. M., & Coros, S. (2019). Puppetmaster: robotic animation of marionettes. *ACM Transactions on Graphics (TOG)*, *38*(4), 1–11.