NaP-Control: Navigating Diffusion Prior for Versatile and Fast Character Control

Chia-Wen Chen, Yan Wu, Korrawe Karunratanakul, and Siyu Tang

ETH Zurich
chiachen@ethz.ch, {yan.wu, korrawe.karunratanakul, siyu.tang}@inf.ethz.ch
https://chiawenchen.github.io/nap-control-project/

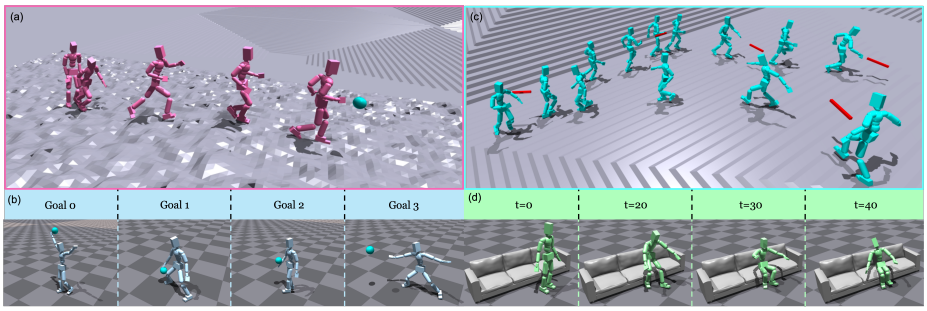

Fig. 1: NaP-Control is a latent noise optimization framework combining reinforcement learning and diffusion-based prior for physics-based character control. We showcase its effectiveness in (a) far goal reaching, (b) agile hand reaching, (c) velocity control, (d) object interaction tasks, as well as its adaptation on uneven terrains.

Abstract. Achieving precise, versatile whole-body character control in physics-based animation remains challenging. Recent diffusion-based policies generate rich and expressive motions but typically rely on gradient-based test-time guidance to satisfy task objectives, which is slow and can reduce robustness. We introduce **NaP-C**ontrol (**Na**vigating Diffusion **P**rior for Versatile and Fast Character **C**ontrol), abbreviated as **NaP**. Our method uses reinforcement learning to manipulate the latent noise of a task-agnostic diffusion policy prior, steering it toward task-specific behaviors for fast, robust control with high motion fidelity. In contrast to methods that rely solely on offline training, NaP interacts with the environment during training to correct motions and optimize task rewards, improving success rates and enabling adaptation to challenging scenarios. By directly predicting task-optimized diffusion noise, NaP eliminates iterative guidance during denoising and enables efficient inference. Experiments show that NaP attains higher success rates and faster inference while preserving natural motion across diverse tasks.

Keywords: Physics-based Character Control · Diffusion Models · Reinforcement Learning

1 Introduction

Physics-based whole-body character control [13, 25, 26, 28, 39, 40, 50, 52, 57, 58] aims to generate physically plausible human motion that remains robust and responsive under a wide range of control inputs and in complex environments. A central challenge is to produce high-quality behaviors that both satisfy physical constraints and adapt efficiently to different task objectives.

A prominent line of work tackles this challenge by learning reusable motion priors from large-scale motion datasets. In particular, VAE-based action priors [25, 39, 57] allow reinforcement learning (RL) to operate in a compact latent space, but their stateless, single-step formulation can limit temporal consistency and overall expressiveness. More recently, diffusion-based policies [13, 21, 50] have advanced physics-based character control by modeling long-horizon motion distributions, yielding smoother and more expressive behaviors. However, these approaches are generally trained purely offline without interacting with the environment during learning, which constrains their robustness for downstream control tasks. To incorporate task objectives, prior work [21, 50] typically employs loss-based guidance at test time, which requires differentiable objectives and iterative optimization during inference. This results in limited applicability and substantial computational overhead. Consequently, how to effectively leverage expressive diffusion policy priors for robust and efficient task adaptation in physics-based character control remains largely an open question.

To address this challenge, we introduce NaP-Control, a framework that adapts a pretrained task-agnostic diffusion policy by learning to navigate its input noise space through RL. Prior work has investigated diffusion-noise optimization for behavioral control, including RL-based steering in robotic manipulation [45] and direct noise optimization for kinematics-based motion synthesis [15]. Our method extends this emerging paradigm to a substantially more challenging setting: physics-based whole-body character control, in which task performance depends on closed-loop interaction with the environment, contact-rich dynamics, and long-horizon stability. In this setting, small perturbations can induce loss of balance, rendering purely offline methods inadequate. We therefore treat the pretrained diffusion policy as an expressive action manifold and train an RL policy to generate task-aware latent noise through direct interaction with the environment. This formulation enables the controller to adapt to environmental feedback, optimize task-specific rewards, and accommodate tasks that lack simple differentiable objective functions, while preserving the naturalness and smoothness encoded by the diffusion prior. Notably, NaP-Control eliminates the need for both extensive reward engineering, which is typically required in reinforcement learning, and iterative gradient-based guidance at inference time, thereby achieving substantially higher computational efficiency than existing diffusion-based control frameworks.

Our contributions are summarized as follows:

(1) We introduce NaP-Control, a reinforcement learning framework that adapts pretrained diffusion-based action priors for task-driven physics-based character control by producing the initial diffusion noise.

(2) We show that navigating the initial diffusion noise space provides an effective way to incorporate environmental feedback into diffusion-based control, overcoming the limitations of purely offline training and expensive test-time guidance while preserving motion fidelity.

(3) We demonstrate that NaP-Control achieves faster inference and higher success rates across diverse locomotion tasks, including challenging traversal on unseen terrains, compared to diffusion-based policies while maintaining natural and human-like whole-body movement in all tasks.

2 Related Work

2.1 Kinematics-based Human Motion Generation

Human motion generation aims to synthesize realistic motion sequences, typically represented as trajectories of joint rotations and global root motion over time. A large body of prior work focuses on kinematics-based generation, which emphasizes visual realism and diversity without explicitly modeling physical dynamics. These methods have achieved strong results in multi-modality conditioned human motion generation [10,17,18,43,48,56,60], and controllable motion editing and synthesis [15, 16, 19, 32].

A central direction in this area is controllable motion synthesis. Existing approaches [14, 15, 32, 46] introduce control through target trajectories, obstacle constraints, or joint-level spatio-temporal specifications. In diffusion-based methods, such control is often incorporated through guidance [5, 16, 37, 53] or noise optimization [15], while other methods allow more fine-grained editing of specific body parts [33] or motion segments [32]. More recent work has also extended kinematics-based generation to online settings [42, 51, 62], where motion is produced autoregressively or in a streaming manner based on motion history and language input, enabling more responsive and continuous generation.

Despite their strong visual quality, kinematics-based methods do not explicitly enforce physical consistency, and thus often suffer from artifacts such as floating, foot skating, ground penetration, and jitter [59]. These limitations make them difficult to deploy in embodied settings and motivate physics-based motion generation, which incorporates dynamics and contact interactions to produce physically feasible behaviors.

2.2 Physics-based Character Control

Early work in physics-based character control trains task-specific controllers using reinforcement learning [6, 23, 28]. To improve motion realism, later methods incorporate motion capture data through imitation and adversarial learning [29, 30, 41]. While effective, these approaches typically require separate policies for different tasks, limiting scalability and generalization. Recent work learns generalized motion tracking controllers or reusable action priors from large-scale motion datasets [21, 26, 39, 40, 47, 57, 58]. For example, PHC [26] and MaskedMimic [39] show that a single controller can track diverse motions and support

multiple tasks. PULSE [25] distills a VAE-based action prior from motion data to enable efficient downstream learning, while MaskedMimic [39] learns a multi-task controller. Nevertheless, VAE-based priors and single-step formulations often limit motion expressiveness and temporal coherence. More recently, diffusion models have significantly improved motion modeling and have been applied to physics-based control [13, 42, 44, 50]. CLoSD [42] integrates a diffusion-driven kinematic generator with a physics-based tracker. Other methods, including UniPhys [50] and DiffuseCLoC [13], distill diffusion policies from offline datasets and adapt them to new tasks using test-time guidance. Yet, these techniques rely on iterative, gradient-based optimization during inference, resulting in slow and sometimes unstable control and limiting adaptation due to the lack of direct environment interaction.

In contrast, we build upon a pretrained diffusion policy and propose an efficient framework to leverage it as a temporally smooth, expressive action prior. By learning to navigate the latent noise space using task rewards and online environment feedback, our method enables robust, goal-directed behavior without costly test-time guidance, while supporting a wide range of downstream tasks with substantially faster inference.

2.3 Improving Diffusion Model with Reinforcement Learning

Reinforcement learning (RL) has recently been explored as an effective way to improve and adapt pretrained models across domains such as image generation [3, 8, 9], language modeling [1, 22], and robotics [34, 45]. The goal is to align generative outputs with downstream objectives, such as human preference and physical feasibility, which are difficult to capture through supervised learning alone. One approach directly fine-tunes diffusion models using reward signals [3, 4, 34, 61]. Examples include RLHF-style methods [4] for improving image generation and DPPO [34] and IDQL [12] for fine-tuning diffusion policies in robotics. However, while effective, joint optimization is often unstable and sample-inefficient. Another direction leverages RL without modifying the pretrained diffusion model by learning an auxiliary policy, such as learning a residual policy [49] on top of the frozen prior, though these methods mainly provide local corrections. More recently, latent noise steering [7, 45] has been proposed, where RL controls the input noise of the diffusion process, and DSRL [45] shows that noise steering can effectively control diffusion-based policies for low-dimensional robotic manipulation. In contrast to prior work that primarily targets low-dimensional or kinematic settings, we scale latent noise steering to high-dimensional, physics-based whole-body control.

3 Whole-Body Control via Diffusion Noise Navigation

To facilitate the generation of natural motions that satisfy task objectives, we propose a framework that utilizes a pre-trained diffusion model as a generative action prior and steers it toward task-specific control. We hypothesize that a

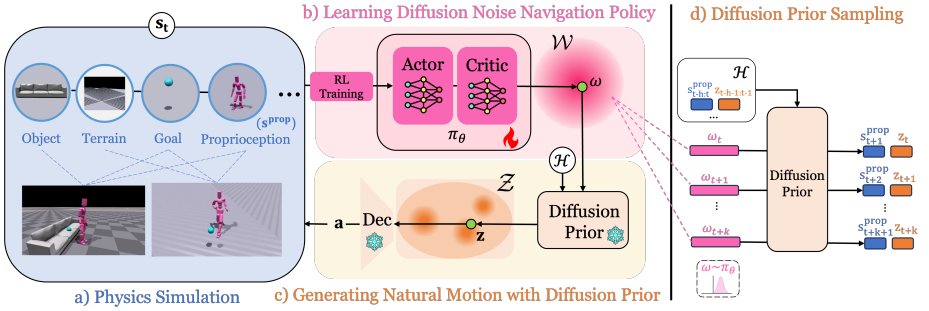

Fig. 2: Framework overview. (a) The RL policy π_θ receives environment and proprioceptive states from the simulation. (b) The actor learns to predict optimal noise $\boldsymbol{\omega} \in \mathcal{W}$ aligned with task goals. (c) These predicted noises are denoised and decoded into executable actions $\mathbf{a}$ via a pretrained diffusion prior and a latent action decoder. Resulting transitions are then used to iteratively optimize the noise navigation policy with RL. (d) As a natural motion manifold, the diffusion prior can map predicted noises to future states $\mathbf{s}^{\text{prop}}$ and latent actions $\mathbf{z}$ conditioned on the trajectory history $\mathcal{H}$.

task-agnostic diffusion policy provides an expressive movement prior. By searching within this prior space, a wide range of downstream tasks can be achieved while preserving motion smoothness and naturalness, which are inherently regularized by the action prior. We therefore pretrain a diffusion-based action prior (Section 3.1) and demonstrate that optimizing the initial noise of the action prior through RL (Section 3.2) allows for the effective learning of task-specific policies (Section 3.3). Fig. 2 provides an overview of the overall framework.

3.1 Diffusion-based Action Prior Pretraining

We first pretrain a diffusion-based behavior generative model that serves as a task-agnostic action prior, grounding policy rollouts within a high-fidelity motion manifold. This prior captures the distribution of natural humanoid behaviors and later acts as a strong regularizer for downstream task learning. We leverage a large-scale state–action dataset constructed by tracking reference human motions from AMASS [27] in a physics simulator. Following UniPhys [50], the diffusion prior is implemented as a causal transformer-based diffusion model that conditions on the history of previous states and actions. Given this context, the model iteratively denoises latent noise to generate a chunk of future actions and predicted states, as shown in Fig. 2(d). We adopt their training paradigm while introducing a more compact state representation that removes redundant kinematic features. Specifically, UniPhys [50] employs an expansive state representation that contains a variety of kinematic features, including joint positions, velocities, and other redundant quantities. In contrast, we retain only the essential dynamic features $\mathbf{S} = (\mathbf{r}_{1:T}, \mathbf{q}_{1:T}, \mathbf{w}_{1:T})$, which include:

- Global Root Trajectory ($\mathbf{r}_{1:T}$): The root trajectory is canonicalized relative to the first-frame coordinate system, where the initial position is at the origin facing the positive y-axis. It includes the root position $\zeta_t \in \mathbb{R}^3$, 6D orientation $\phi_t \in \mathbb{R}^6$, linear velocity $\dot{\zeta}_t \in \mathbb{R}^3$, and angular velocity $\dot{\phi}_t \in \mathbb{R}^3$.
- Local Joint Features: Joint-level data is canonicalized to a per-frame local coordinate system centered at the pelvis projection on the ground. This includes 6D joint rotations $\mathbf{q}_t \in \mathbb{R}^{J\times 6}$, and angular velocities $\mathbf{w}_t \in \mathbb{R}^{J\times 3}$, and exclude local joint positions $\mathbf{p}_t \in \mathbb{R}^{J\times 3}$ and velocities $\mathbf{v}_t \in \mathbb{R}^{J\times 3}$.

This compact representation maintains motion fidelity and robustness while lowering the dimensionality of the learning space. Consequently, both diffusion-prior pretraining and the following reinforcement learning are more efficient.

We model the joint state-action distribution with the diffusion prior following previous works [13,50]. To manage the high dimensionality of the raw action space $\mathcal{A}$, we use a representation from PULSE [25] to first encode the raw action into a compact latent action space $\mathcal{Z}$. The diffusion prior therefore models the trajectory of state–latent pairs $\mathbf{x}_t = (\mathbf{s}_t, \mathbf{z}_t)$. During inference, the model conditions on the historical trajectory $\mathcal{H}$ and iteratively denoises a sequence of future latent actions, which are then decoded into executable control signals via the pretrained action decoder $\mathbf{a_t} = \mathcal{D}(\mathbf{s}_t, \mathbf{z}_t)$.

3.2 Navigating Diffusion Noise for Whole-Body Control

To navigate the manifold of the pre-trained diffusion prior and satisfy downstream task objectives, we learn a latent noise navigation policy via RL to select the initial diffuse noise. This navigation mechanism is critical, as the initial noise determines the specific trajectory the diffusion process takes through the learned motion space [11, 15]. By treating the noise space as a controllable search manifold, we can effectively guide the generative process toward physically plausible and task-specific behaviors.

Navigation Mechanism in Natural Motion Prior. As illustrated in Fig. 2a, the latent noise navigation policy π_θ processes the observation $\mathbf{s}_t = \{\mathbf{s}^{\text{prop}}, \mathbf{s}^{\text{env}}\}$, where $\mathbf{s}^{\text{prop}}$ denotes the character's current proprioceptive state and $\mathbf{s}^{\text{env}}$ encodes task-specific environment state. Unlike standard controllers that map states directly to raw joint actions $\pi(\mathbf{a} \mid \mathbf{s})$, our navigation policy π_θ, following the formulation of DSRL [45], operates on the latent noise manifold of a pretrained diffusion prior. To preserve the diffusion prior's sequential generation and temporal consistency while maintaining a low-dimensional search space for efficient RL exploration, we duplicate the latent noise $\boldsymbol{\omega}_t$ into a k-step noise chunk [45]. The predicted noise chunk $\boldsymbol{\omega}_{t:t+k-1}$ is further denoised into a task-aligned latent action chunk $\mathbf{z}_{t:t+k-1}$ using the DDIM-ODE solver [38]:

$$\boldsymbol{\omega}_t = \pi_\theta(\mathbf{s}_t) \tag{1}$$

$$\mathbf{z}_{t:t+k-1} = \mathcal{M}(\text{repeat}(\boldsymbol{\omega}_t, k)|\mathcal{H}) \tag{2}$$

The latent actions are then decoded to joint-level raw actions $\mathbf{a}_{t:t+k-1}$ via the frozen action decoder $\mathcal{D}$ [25]. The joint-level actions are executed open-loop, i.e., without per-step feedback, within the physics simulator, which returns the subsequent proprioceptive $\mathbf{s}_{t+1:t+k}$ and environmental observations. We encapsulate both the diffusion prior as well as the latent action decoder as part of the Markov Decision Process (MDP) environment for our latent noise navigation policy, as shown in Fig. 2(c). This architecture is crucial for grounding the final actions within a natural, human-like motion manifold. By doing so, it ensures physical plausibility and prevents the character from drifting into out-of-distribution states, a common failure mode in traditional test-time guidance methods.

Algorithm 1 Latent Noise Navigation Policy Training

Input: Frozen diffusion prior $\mathcal{M}$, frozen decoder $\mathcal{D}$, physics simulator $\mathcal{SIM}$, reward function $r(\cdot)$, terrain heightmap H (optional), action chunk size k, initial proprioceptive state $\mathbf{s}^{prop}$ and environment states $\mathbf{s}^{env}$, horizon length T
Output: Noise optimizing policy π_θ
Initialize policy π_θ, experience buffer $\mathcal{B}$, and diffusion history buffer $\mathcal{H}$
for epoch = 1, 2, ..., N **do**
  Reset environment; obtain $\mathbf{s}^{prop}$ and $\mathbf{s}^{env}$ from $\mathcal{SIM}$
  Reset diffusion history buffer $\mathcal{H}$
  for timestep $t = 0, k, 2k, \ldots, T-k$ **do**
    $\mathbf{s}_t \leftarrow \{\mathbf{s}_t^{prop}, \mathbf{s}_t^{env}, \mathcal{H}\}$
    Sample latent noise $\boldsymbol{\omega} \sim \pi_{\theta_{old}}(\boldsymbol{\omega} \mid \mathbf{s}_t)$
    Denoise to latent action chunk: $\mathbf{z}_{t:t+k-1} = \mathcal{M}(\text{repeat}(\boldsymbol{\omega}, k) \mid \mathcal{H})$
    $R_{chunk} \leftarrow 0$
    for $i = 0$ **to** $k-1$ **do**
      $\mathbf{a}_{t+i} = \mathcal{D}(\mathbf{s}_{t+i}, \mathbf{z}_{t+i})$
      Execute $\mathbf{a}_{t+i}$ in $\mathcal{SIM}$, observe $\mathbf{s}_{t+i+1}$
      Compute reward $r_{t+i+1} = r(\mathbf{s}_{t+i+1})$
      Update $\mathcal{H} \leftarrow \{\mathbf{s}_{t+i+1}, \mathbf{z}_{t+i}\}$
      $R_{chunk} \leftarrow R_{chunk} + \gamma^i r_{t+i+1}$
    end for
    Store transition $(\mathbf{s}_t, \boldsymbol{\omega}, R_{chunk}, \mathbf{s}_{t+k})$ in $\mathcal{B}$
  end for
  Compute advantage estimates $\hat{A}_t$ based on $\mathcal{B}$, for $t \in \{0, k, 2k, \ldots, T-k\}$
  Optimize surrogate loss L wrt θ (m miniepochs, minibatch size b)
  $\pi_{\theta_{old}} \leftarrow \pi_\theta$
end for
return π_θ

Latent Noise Navigation Policy Training. To learn the latent noise navigation policy, our framework employs on-policy Reinforcement Learning (RL) via Proximal Policy Optimization (PPO) [36], shown in Fig. 2b. During training, the policy iteratively shifts from producing stochastic noise with low task-efficacy toward generating noise distributions optimized for task fulfillment. We optimize

actor and critic networks using the PPO clipped objective [36] and Generalized Advantage Estimation (GAE) [35] to ensure training stability. We also incorporate a bounding loss on the policy's output, ensuring the generated noise remains within a stable value range. We treat the entire action chunk as a single micro-step in an MDP. Accordingly, rewards r_i are calculated at each simulation step, but accumulated through a discount factor γ to form the total chunk reward. The resulting transition tuple $(\mathbf{s}_t, \boldsymbol{\omega}, \sum_{i=1}^{k} \gamma^{i-1} r_{t+i}, \mathbf{s}_{t+k})$ is stored in the experience buffer for policy updates and $\mathbf{s}_{t+1:t+k}$ is saved to diffusion prior's history buffer $\mathcal{H}$ to provide temporal consistency for the next roll out. The complete training algorithm is shown in Algorithm 1.

In this manner, our framework uses the diffusion prior as a task-agnostic natural motion manifold, exploiting its expressive motion space to generate high-fidelity actions that both preserve motion realism and optimize downstream task rewards. By directly interacting with the environment during training, the framework adapts the pretrained prior to handle scenarios beyond the original motion distribution, including uneven terrain and object interaction. By replacing the standard stochastic Gaussian sampling $\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ used in traditional diffusion guidance with the deterministic noise output of our trained policy π_θ, we eliminates the reliance on computationally expensive test-time guidance, enabling more efficient goal-conditioned task execution. In summary, the latent noise navigation policy together with the diffusion prior effectively bridges high-level task constraints with the low-level natural motion manifold, enabling robust, fast and versatile whole-body control.

3.3 Applications

We demonstrate that our proposed framework is effective across various locomotion tasks while maintaining the intrinsic naturalness of the underlying motion manifold. Moreover, it exhibits robust performance in extended and complex scenarios, including object interaction and uneven terrain traversal, which are typically challenging for previous guidance-based diffusion policies [13, 50].

Far Goal Reaching. The agent is tasked with reaching a target position p_{goal}, as illustrated in Figure 1(a). The reward consists of three terms: (1) a location reward that drives the agent toward the goal; (2) an orientation reward that encourages the agent to face the goal during approach; and (3) a stability penalty that promotes controlled, steady arrival. The stability term is crucial for suppressing unnatural orbiting behavior, where the agent circles around the goal to preserve momentum instead of settling at p_{goal}. More concretely,

$$R_{\text{reach}} = (1 - \lambda_1 - \lambda_2) \cdot R_{\text{location}} + \lambda_1 \cdot R_{\text{orientation}} + \lambda_2 \cdot R_{\text{stability}} \tag{3}$$

$$R_{\text{location}} = \exp(-\alpha_1 \cdot \|p_{\text{goal}} - p_{\text{joint}}\|_2) \tag{4}$$

$$R_{\text{orientation}} = \begin{cases} 1.0 & \text{if } \|p_{\text{goal}} - p_{\text{joint}}\|_2 \leq 0.3\,\text{m} \\ \exp\left(\alpha_2 \cdot (\mathbf{d}_{\text{fwd}} \cdot \mathbf{d}_{\text{goal}} - 1.0)\right) & \text{otherwise} \end{cases} \tag{5}$$

$$R_{\text{stability}} = \begin{cases} -\beta_1 \cdot (c_1 v + c_2 \omega) & \text{if } \|p_{\text{goal}} - p_{\text{joint}}\|_2 < 1.2\,\text{m} \\ -\beta_2 \cdot (c_1 v + c_2 \omega) & \text{if } 1.2\,\text{m} \leq \|p_{\text{goal}} - p_{\text{joint}}\|_2 < 6.0\,\text{m} \\ 0 & \text{if } \|p_{\text{goal}} - p_{\text{joint}}\|_2 \geq 6.0\,\text{m} \end{cases} \tag{6}$$

Agile Hand Reaching. We further show that our framework is well-suited for tasks that demand high agility and precise end-effector control. We design an agile hand-reaching task (Fig. 1(b)) in which the agent must move its right hand to targets distributed throughout its reachable workspace. The task is solved using only a simple location reward (Eq. 4), without additional reward shaping or engineering. Despite this minimal design, the character reaches targets naturally and robustly, exhibiting seamless whole-body coordination such as stopping, squatting, and side-stepping when needed.

Velocity Control. We consider a velocity control task where the agent is required to track time-varying target speeds and headings. We define the character velocity $\mathbf{v}_{\text{joint}}$ as the linear velocity of the pelvis joint projected onto the horizontal plane. The velocity reward R_{velocity} is formulated as a weighted sum of a velocity tracking term and an orientation alignment term;

$$R_{\text{velocity}} = (1-\lambda) \cdot \exp\left(-\alpha_3 \cdot \|\mathbf{v}_{\text{target}} - \mathbf{v}_{\text{joint}}\|^2\right) + \lambda \cdot \frac{\mathbf{d}_{\text{fwd}} \cdot \mathbf{d}_{\text{target}} + 1}{2} \tag{7}$$

where $\mathbf{d}_{\text{fwd}}$ and $\mathbf{d}_{\text{target}}$ represent the current and target heading vectors, respectively. This formulation ensures that the character not only reaches the target speed but also aligns its body orientation with the direction of travel for more natural movement, as demonstrated in Fig. 1(c).

Object Interaction. To evaluate our framework's ability to handle contact-rich interactions, we introduce a sofa-sitting task, shown in Fig. 1(d). The agent is required to lower its center of mass onto a target seating region on the sofa, whose position and yaw are randomized at initialization to encourage generalization. The reward follows the same formulation as R_{location}, penalizing the distance between the character's pelvis and the target seating region.

Adaptation on Uneven Terrain. We hypothesize that, although the pretrained diffusion prior may fail to maintain balance on uneven terrain, its motion manifold still contains latent primitives, *e.g.*, high-stepping and balance recovery, that can be leveraged for locomotion in challenging environments. To test

this, we evaluate our policies on uneven terrains, which are unseen in the diffusion prior's training data. Unlike loss-based test-time guidance, which would require an explicit analytical model of complex terrain geometry, our approach simply augments the policy observation with a terrain height map. The height map encodes terrain elevations within a 4×4 meter square centered on the agent's head at a resolution of 12.5 cm.

To facilitate stable locomotion on irregular surfaces, we employ curriculum learning [2] for both far goal reaching and agile hand reaching (Fig. 1(a),(c)). Training starts on relatively flat terrain with nearby targets. Once the mean reward exceeds a threshold, we progressively increase both terrain difficulty, *e.g.*, steeper stairs, rougher slopes, discrete blocks, and target distance. Please refer to the supplementary materials for more details.

4 Experiments and Results

4.1 Experiment Setup

Implementation Details. When pre-training the diffusion prior, we follow the training paradigm of UniPhys [50], utilizing a 12-layer causal Transformer decoder with a 768-dimensional hidden state and 8 attention heads. The model takes 224-dimensional compact state representations alongside 32-dimensional latent action embeddings as input. When training the noise navigation policy with PPO, we freeze the pretrained diffusion prior and use 5-step DDIM denoising. To balance reactivity and efficiency, the action chunk size k is dynamically adjusted based on task demands: $k = 4$ is utilized for uneven terrain and agile reaching tasks to enhance near closed-loop control, while $k = 8$ is used for flat-ground scenarios to maximize inference speed. Within each chunk, the pipeline outputs target joint positions $\mathbf{a}_t \in \mathbb{R}^{J\times 3}$, which are converted to torques via a proportional-derivative (PD) controller. These torques drive the transition to the next state $\mathbf{s}_{t+1} = \mathcal{SIM}(\mathbf{s}_t, \mathbf{a}_t)$ within the Isaac Gym simulator [20]. The simulation environment employs a 30Hz control frequency and an SMPL-like character [24] with 24 rotational joints. We adopt an early termination criterion if the pelvis height drops below 0.15 m [28]. More PPO training details and task-specific reward settings are available in the supplementary material.

Baselines. We evaluate our method against four representative physics-based character control baselines. (1) UniPhys [50], a unified diffusion-based policy that relies on gradient-based test-time guidance for task-specific control; (2) PULSE [25], a per-task RL policy utilizing a VAE prior; (3) CLoSD [42], a hierarchical framework that decouples control into a kinematic diffusion planner and an RL-based low-level tracker; and (4) MaskedMimic [39], a unified RL controller that distills a multi-task policy using a masked cVAE. In addition to these general controllers, we evaluate against three representative task-specific RL baselines utilizing adversarial motion priors for the far goal-reaching task:(5) AMP [31], (6) CML [54], and (7) AdaptNet [55]. We used the officially released

Table 1: Comparison with task-specific RL baselines on the far-goal reaching task.

Type	Methods	Jerk (m/s^3) ↓	Success Rate ↑	Samples ↓
Flat	AMP [31]	2387	0.967	1200M
	CML [54]	2185	0.963	400M
	NaP(Ours)	**1266**	**0.984**	327M
Terrain	AdaptNet [55]	2444	**0.868**	408M
	NaP(Ours)	**1788**	0.860	900M

models for all baselines, except for PULSE and AMP, which we trained with the same reward as our method.

Evaluation Metrics. For all tasks, we measure task success rates and error metrics across 1000 episodes to evaluate the policy performance. We further assess physical plausibility by quantifying motion smoothness via jerk, the time derivative of acceleration. Fallen cases are excluded from these plausibility measurements. We also benchmark the inference efficiency (frames per second, FPS) against the diffusion-based baseline. To evaluate sample efficiency, we report the total number of Markov Decision Process (MDP) steps in Table 1, where each step corresponds to an action segment spanning 4 frames for terrain, and 8 frames for flat ground. Sample counts for AMP are obtained from our own training runs, while metrics for CML and AdaptNet are sourced directly from their original papers. More qualitative visual results can be found in the supplementary video.

4.2 Far Goal Reaching

This task evaluates long-horizon stability through far-goal reaching. During training, pelvis goals are set within a 0.9–0.92 m height and a 6 m horizontal radius. During evaluation, goals are fixed at a 5 m distance and uniformly sampled within a $\pm 45^\circ$ sector relative to the character's forward heading. Success requires the pelvis to stay within 0.3 m of the goal for 0.5 s.

As presented in Table 2 and Table 3, compared with the diffusion-based UniPhys, which depends on test-time guidance to reach the goal, NaP achieves both substantially faster inference and higher success: 22.5 FPS vs. 2.9 FPS and 98.4% vs. 81.9%, while still preserving motion smoothness and naturalness. Against RL-based baselines (CLoSD and PULSE), NaP attains a similarly high success rate but produces significantly smoother and more natural motions. In particular, PULSE exhibits noticeable high-frequency jitter and less fluid transitions, whereas CLoSD often yields abrupt turning behaviors and severe jittering, especially in the hand and foot joints. MaskedMimic, in contrast, fails to reach distant goals robustly. Similarly, compared to task-specific adversarial baselines, NaP attains higher motion naturalness than AMP as shown in Table 1 and delivers superior motion quality over CML and AdaptNet while maintaining a

Table 2: Inference efficiency comparison with diffusion-based baseline.

Inference Speed (FPS) ↑	Far Goal Reaching	Velocity Control
UniPhys	2.9	5.2
NaP-Control (Ours)	22.5	24.4

Table 3: Flat-ground motion naturalness and success rate comparison.

Methods	Far Goal Reaching		Agile Hand Reaching		Velocity Control			Object Interaction	
	Jerk ↓ (m/s^3)	Success Rate ↑	Jerk ↓ (m/s^3)	Success Rate ↑	Jerk ↓ (m/s^3)	Velocity Error ↓ (m/s)	Success Rate ↑	Jerk ↓ (m/s^3)	Success Rate ↑
PULSE [25]	2594	**0.998**	3631	**0.998**	2273	**0.0455**	**0.996**	1373	**0.999**
MaskedMimic [39]	3564	0.304	2750	0.725	2420	0.0678	0.355	1310	0.298
CLoSD [42]	1876	**0.998**	–	–	–	–	–	1505	0.862
UniPhys [50]	**528**	0.819	–	–	**401**	0.7971	0.921	–	–
NaP-Control(Ours)	1266	0.984	**1146**	0.991	752	0.1753	0.990	**491**	0.979

comparable success rate. We also compare against a transformer-based task-specific RL baseline on a path-following task. Please refer to the supplementary material for details.

These results demonstrate that navigating latent noise selection in a pre-trained diffusion-based action prior using task rewards and environmental feedback enables robust, goal-directed behavior while preserving motion realism. Importantly, NaP removes the need for expensive test-time optimization, as the policy directly learns to generate physically feasible and task-successful actions through efficient noise navigation. Additional qualitative comparisons are provided in the supplementary video.

4.3 Agile Hand Goal Reaching

This task evaluates motion agility and reactivity through right-hand reaching with frequently changing targets sampled from all directions. During evaluation, targets are placed within a radius of 1 m, with heights ranging from 0.4 to 2.0 m. The task requires rapid postural adaptation while maintaining temporal smoothness and stability. A trial is considered successful if the hand remains within 0.3 m of the target for at least 0.2 s.

UniPhys and CLoSD fail to complete this task due to insufficient fine-grained control, while MaskedMimic achieves a relatively low success rate. Consistent with the far-goal reaching results, NaP achieves a success rate comparable to PULSE, but with substantially reduced motion jitter (jerk 1146 m/s^3 vs. 3631 m/s^3) and noticeably smoother transitions. Compared to far goal reaching, the advantage in motion naturalness becomes even more pronounced in this setting, as the task demands rapid directional and height changes that strongly challenge the temporal coherence of conventional RL-based policies.

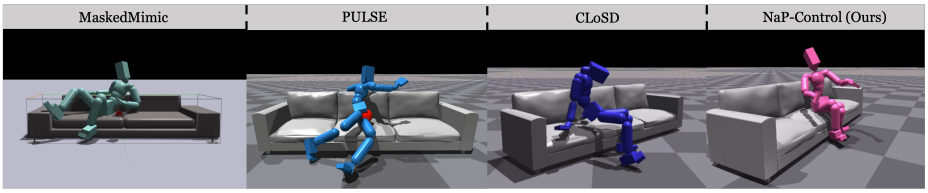

Fig. 3: Qualitative Comparison of Object Interaction Task.

4.4 Velocity Control

For the velocity control task, the direction and speed of the target velocity are randomly sampled within $3m/s$ for each episode. We allow a 90-frame (3s) transition period to reach the target velocity before calculating the mean tracking error, while motion smoothness is quantified over the entire 200-frame (7s) non-falling episodes to capture the quality of the initial acceleration and facing-direction transition phase. The success rate is defined as the ratio of episodes in which the character completes the motion without falling to the total number of episodes. CLoSD does not provide velocity control, and MaskedMimic fails the task with a low success rate. PULSE achieves a low velocity error, but produces unnatural running poses with a high jerk. In contrast, UniPhys produces much smoother motion with a jerk of 401 m/s^3, but at the cost of a significant velocity error of 0.7971 m/s. NaP effectively bridges this gap, generating smooth motion with a jerk of 752 m/s^3, which is comparable to the performance of UniPhys, while achieving a velocity error of 0.1753 m/s.

4.5 Object Interaction

We evaluate the agent's ability to reach a target sitting position on a sofa. Success is achieved when the pelvis is within 0.15m horizontally and 0.1m vertically of the target height. Although our diffusion prior was not trained on human-object interaction data, NaP can identify the most suitable motion for the sitting task while maintaining natural sitting motions. In contrast, CLoSD and MaskedMimic have a low task success rate; UniPhys cannot perform the task; PULSE, without a specific reward for human-like motion, produces highly unnatural behaviors to satisfy the goal, such as severe upper-body twisting (see Fig. 3). This comparison underscores that NaP inherently preserves the manifold of natural human movement, even when applied to object interaction scenarios for which no explicit data exist in the training of the diffusion prior.

4.6 Performance on Uneven Terrain

By incorporating terrain height maps, we extend our framework to uneven terrains, demonstrating strong generalization to previously unseen environments. We note that diffusion-based baselines such as UniPhys and CLoSD are trained

Table 4: Uneven-terrain motion naturalness and success rate comparison. While MaskedMimic minimizes velocity error, it suffers from a lower success rate.

Methods	Far Goal Reaching		Agile Hand Reaching		Velocity Control		
	Jerk $(m/s^3)\downarrow$	Success Rate ↑	Jerk $(m/s^3)\downarrow$	Success Rate ↑	Jerk $(m/s^3)\downarrow$	Velocity Error $(m/s)\downarrow$	Success Rate ↑
PULSE [25]	2830	**0.966**	3976	**0.980**	2397	<u>0.0569</u>	**0.995**
MaskedMimic [39]	3566	0.319	2723	0.676	2262	**0.0556**	0.292
NaP-Control(Ours)	**1788**	<u>0.860</u>	**1520**	<u>0.909</u>	**1358**	0.1654	<u>0.950</u>

and evaluated on flat-ground settings in their original formulations, and therefore cannot be directly deployed on uneven terrains. Consequently, we restrict our benchmarking to PULSE and MaskedMimic. Consistent with our flat-ground results, NaP achieves goal-reaching success rates and velocity tracking errors comparable to PULSE, and significantly outperforms MaskedMimic, while producing substantially smoother and more human-like motion. These results highlight that navigating latent noise selection within a pretrained diffusion-based action prior enables robust adaptation to new environments, allowing our approach to extend naturally to more challenging and previously unseen scenarios. Quantitative comparisons are summarized in Table 4.

4.7 Ablation Studies

Motion State Representation. We adopt a compact yet informative state representation that includes global root trajectories, joint rotations, and angular velocities, while excluding local joint positions and linear velocities. This design reduces the effective search space without sacrificing key motion cues, balancing RL exploration efficiency and control stability, as evidenced by our flat-ground far goal-reaching results (Fig. 4(a)). Using this compact representation yields more efficient exploration than the redundant full state used in UniPhys and substantially outperforms a minimal state that retains only root-related features.

Action Chunking Size. While action chunks of 4–8 steps generally balance temporal consistency and reactive control (Fig. 4(b)), task difficulty dictates the ideal granularity. As shown in Fig. 4(c), smaller chunk sizes favor stability on uneven terrain and achieving precision in agile hand-reaching tasks.

Noise Navigation Space. While task execution only requires action output, we find that navigating both state and action noise spaces significantly outperforms navigating action noise space alone. As evaluated on flat-ground agile hand reaching (Fig. 4(d)), this joint strategy provides the RL policy with finer control over the diffusion prior's generative manifold, resulting in more robust and physically plausible task completion.

ODE-Solver Step and Inference Speed. We evaluate the inference speed across varying diffusion denoising steps for the velocity control task on flat

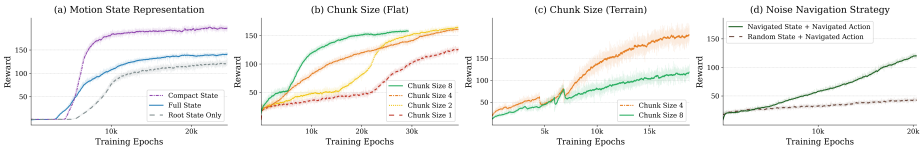

Fig. 4: Ablation studies. (a) Effect of state representation on flat-ground far goal reaching. (b–c) Effect of action chunk size k for agile hand reaching on flat ground (b) and uneven terrain (c). (d) Comparison of joint state-action noise optimizing versus action-only noise optimizing for agile hand reaching.

Steps	FPS
3	38
4	37
5	35
10	31

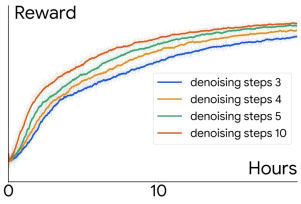

Fig. 5: Training convergence of the velocity control task on flat ground with different diffusion denoising steps.

ground. 5 denoising steps are a practical trade-off between training and inference efficiency as shown in Fig. 5. Inference can be accelerated by reducing the number of denoising steps, at the cost of increased training effort.

5 Conclusion

We presented NaP-Control, a framework for fast and robust whole-body character control that navigates the latent noise of a task-agnostic diffusion policy toward task-specific behaviors. By training a compact RL agent to directly predict the diffusion noise that leads to high-reward actions, NaP preserves the motion quality of the prior while eliminating both the computational overhead and fragility of gradient-based control at inference. A key property of NaP is that it learns by interacting directly with the environment, enabling it to generate motion under complex dynamics and handle non-differentiable task objectives, which enhances robustness in difficult conditions. Consequently, NaP achieves higher success rates with low-latency inference, making diffusion-based control viable for real-time use. We believe NaP provides a foundation for generating natural, expressive motion in diverse tasks that require long-horizon stability and precise responsiveness, including whole-body loco-manipulation, contact-intensive interactions, and dynamic multi-agent settings.

References

1. Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., et al.: Training a helpful and harmless assistant with reinforcement learning from human feedback. arXiv preprint arXiv:2204.05862 (2022)
2. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th annual international conference on machine learning. pp. 41–48 (2009)
3. Black, K., Janner, M., Du, Y., Kostrikov, I., Levine, S.: Training diffusion models with reinforcement learning. arXiv preprint arXiv:2305.13301 (2023)
4. Christiano, P.F., Leike, J., Brown, T., Martic, M., Legg, S., Amodei, D.: Deep reinforcement learning from human preferences. Advances in neural information processing systems **30** (2017)
5. Dai, W., Chen, L.H., Wang, J., Liu, J., Dai, B., Tang, Y.: Motionlcm: Real-time controllable motion generation via latent consistency model. In: European Conference on Computer Vision. pp. 390–408. Springer (2024)
6. De Lasa, M., Mordatch, I., Hertzmann, A.: Feature-based locomotion controllers. ACM transactions on graphics (TOG) **29**(4), 1–10 (2010)
7. Eyring, L., Karthik, S., Roth, K., Dosovitskiy, A., Akata, Z.: Reno: Enhancing one-step text-to-image models through reward-based noise optimization. Advances in Neural Information Processing Systems **37**, 125487–125519 (2024)
8. Fan, Y., Lee, K.: Optimizing ddpm sampling with shortcut fine-tuning. arXiv preprint arXiv:2301.13362 (2023)
9. Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., Lee, K.: Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. Advances in Neural Information Processing Systems **36**, 79858–79885 (2023)
10. Gat, I., Raab, S., Tevet, G., Reshef, Y., Bermano, A.H., Cohen-Or, D.: Anytop: Character animation diffusion with any topology. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers. pp. 1–10 (2025)
11. Guo, X., Liu, J., Cui, M., Li, J., Yang, H., Huang, D.: Initno: Boosting text-to-image diffusion models via initial noise optimization. In: CVPR (2024)
12. Hansen-Estruch, P., Kostrikov, I., Janner, M., Kuba, J.G., Levine, S.: Idql: Implicit q-learning as an actor-critic method with diffusion policies. arXiv preprint arXiv:2304.10573 (2023)
13. Huang, X., Truong, T., Zhang, Y., Yu, F., Sleiman, J.P., Hodgins, J., Sreenath, K., Farshidian, F.: Diffuse-cloc: Guided diffusion for physics-based character look-ahead control. ACM Transactions on Graphics (TOG) **44**(4), 1–12 (2025)
14. Huang, Y., Wan, W., Yang, Y., Callison-Burch, C., Yatskar, M., Liu, L.: Como: Controllable motion generation through language guided pose code editing. In: European Conference on Computer Vision. pp. 180–196. Springer (2024)
15. Karunratanakul, K., Preechakul, K., Aksan, E., Beeler, T., Suwajanakorn, S., Tang, S.: Optimizing diffusion noise can serve as universal motion priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1334–1345 (2024)
16. Karunratanakul, K., Preechakul, K., Suwajanakorn, S., Tang, S.: Guided motion diffusion for controllable human motion synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2151–2162 (2023)

17. Khani, A., Rampini, A., Atherton, E., Roy, B.: Unimogen: Universal motion generation. arXiv preprint arXiv:2505.21837 (2025)
18. Li, J., Cao, J., Zhang, H., Rempe, D., Kautz, J., Iqbal, U., Yuan, Y.: Genmo: A generalist model for human motion. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11766–11776 (2025)
19. Li, Z., Cheng, K., Ghosh, A., Bhattacharya, U., Gui, L., Bera, A.: Simmotionedit: Text-based human motion editing with motion similarity prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27827–27837 (2025)
20. Liang, J., Makoviychuk, V., Handa, A., Chentanez, N., Macklin, M., Fox, D.: Gpu-accelerated robotic simulation for distributed reinforcement learning. In: Conference on Robot Learning. pp. 270–282. PMLR (2018), https://api.semanticscholar.org/CorpusID:53084610
21. Liao, Q., Truong, T.E., Huang, X., Tevet, G., Sreenath, K., Liu, C.K.: Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion. arXiv preprint arXiv:2508.08241 (2025)
22. Liu, H., Sferrazza, C., Abbeel, P.: Chain of hindsight aligns language models with feedback. In: The Twelfth International Conference on Learning Representations
23. Liu, L., Yin, K., Van de Panne, M., Shao, T., Xu, W.: Sampling-based contact-rich motion control. In: ACM SIGGRAPH 2010 papers, pp. 1–10 (2010)
24. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graphics (Proc. SIGGRAPH Asia) **34**(6), 248:1–248:16 (Oct 2015)
25. Luo, Z., Cao, J., Merel, J., Winkler, A., Huang, J., Kitani, K.M., Xu, W.: Universal humanoid motion representations for physics-based control. In: The Twelfth International Conference on Learning Representations (2024), https://openreview.net/forum?id=OrOd8PxOO2
26. Luo, Z., Cao, J., Winkler, A.W., Kitani, K., Xu, W.: Perpetual humanoid control for real-time simulated avatars. In: International Conference on Computer Vision (ICCV) (2023)
27. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G., Black, M.J.: AMASS: Archive of motion capture as surface shapes. In: International Conference on Computer Vision. pp. 5442–5451 (Oct 2019)
28. Peng, X.B., Abbeel, P., Levine, S., van de Panne, M.: Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Trans. Graph. **37**(4), 143:1–143:14 (Jul 2018). https://doi.org/10.1145/3197517.3201311, http://doi.acm.org/10.1145/3197517.3201311
29. Peng, X.B., Guo, Y., Halper, L., Levine, S., Fidler, S.: Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. ACM Transactions On Graphics (TOG) **41**(4), 1–17 (2022)
30. Peng, X.B., Ma, Z., Abbeel, P., Levine, S., Kanazawa, A.: Amp: Adversarial motion priors for stylized physics-based character control. ACM Transactions on Graphics (ToG) **40**(4), 1–20 (2021)
31. Peng, X.B., Ma, Z., Abbeel, P., Levine, S., Kanazawa, A.: Amp: Adversarial motion priors for stylized physics-based character control. ACM Trans. Graph. **40**(4) (Jul 2021). https://doi.org/10.1145/3450626.3459670, http://doi.acm.org/10.1145/3450626.3459670
32. Pinyoanuntapong, E., Saleem, M., Karunratanakul, K., Wang, P., Xue, H., Chen, C., Guo, C., Cao, J., Ren, J., Tulyakov, S.: Maskcontrol: Spatio-temporal control for masked motion synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9955–9965 (2025)

33. Pinyoanuntapong, E., Wang, P., Lee, M., Chen, C.: Mmm: Generative masked motion model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1546–1555 (2024)
34. Ren, A.Z., Lidard, J., Ankile, L.L., Simeonov, A., Agrawal, P., Majumdar, A., Burchfiel, B., Dai, H., Simchowitz, M.: Diffusion policy policy optimization. arXiv preprint arXiv:2409.00588 (2024)
35. Schulman, J., Moritz, P., Levine, S., Jordan, M., Abbeel, P.: High-dimensional continuous control using generalized advantage estimation. arXiv preprint arXiv:1506.02438 (2015)
36. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms (2017), https://arxiv.org/abs/1707.06347
37. Shi, Y., Wang, J., Jiang, X., Lin, B., Dai, B., Peng, X.B.: Interactive character control with auto-regressive motion diffusion models. ACM Trans. Graph. **43**(4) (Jul 2024). https://doi.org/10.1145/3658140, https://doi.org/10.1145/3658140
38. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (ICLR) (2021)
39. Tessler, C., Guo, Y., Nabati, O., Chechik, G., Peng, X.B.: Maskedmimic: Unified physics-based character control through masked motion inpainting. ACM Transactions on Graphics (TOG) **43**(6), 1–21 (2024)
40. Tessler, C., Jiang, Y., Coumans, E., Luo, Z., Chechik, G., Peng, X.B.: Maskedmanipulator: Versatile whole-body control for loco-manipulation. arXiv preprint arXiv:2505.19086 (2025)
41. Tessler, C., Kasten, Y., Guo, Y., Mannor, S., Chechik, G., Peng, X.B.: Calm: Conditional adversarial latent models for directable virtual characters. In: ACM SIGGRAPH 2023 Conference Proceedings. pp. 1–9 (2023)
42. Tevet, G., Raab, S., Cohan, S., Reda, D., Luo, Z., Peng, X.B., Bermano, A.H., van de Panne, M.: CLoSD: Closing the loop between simulation and diffusion for multi-task character control. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=pZISppZSTv
43. Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-or, D., Bermano, A.H.: Human motion diffusion model. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=SJ1kSyO2jwu
44. Truong, T.E., Piseno, M., Xie, Z., Liu, K.: Pdp: Physics-based character animation via diffusion policy. In: SIGGRAPH Asia 2024 Conference Papers. pp. 1–10 (2024)
45. Wagenmaker, A., Nakamoto, M., Zhang, Y., Park, S., Yagoub, W., Nagabandi, A., Gupta, A., Levine, S.: Steering your diffusion policy with latent space reinforcement learning. Conference on Robot Learning (2025)
46. Wan, W., Dou, Z., Komura, T., Wang, W., Jayaraman, D., Liu, L.: Tlcontrol: Trajectory and language control for human motion synthesis. arXiv preprint arXiv:2311.17135 (2023)
47. Wang, J., Luo, Z., Yuan, Y., Li, Y., Dai, B.: Pacer+: On-demand pedestrian animation controller in driving scenarios. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 718–728 (2024)
48. Wang, Y., Li, M., Liu, J., Leng, Z., Li, F.W., Zhang, Z., Liang, X.: Fg-t2m++: Llms-augmented fine-grained text driven human motion generation. International Journal of Computer Vision **133**(7), 4277–4293 (2025)
49. Wang, Y., Jiang, H., Yao, S., Ding, Z., Lu, Z.: Sentinel: A fully end-to-end language-action model for humanoid whole body control. arXiv preprint arXiv:2511.19236 (2025)

50. Wu, Y., Karunratanakul, K., Luo, Z., Tang, S.: Uniphys: Unified planner and controller with diffusion for flexible physics-based character control. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2025)
51. Xiao, L., Lu, S., Pi, H., Fan, K., Pan, L., Zhou, Y., Feng, Z., Zhou, X., Peng, S., Wang, J.: Motionstreamer: Streaming motion generation via diffusion-based autoregressive model in causal latent space. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10086–10096 (October 2025)
52. Xiao, Z., Wang, T., Wang, J., Cao, J., Zhang, W., Dai, B., Lin, D., Pang, J.: Unified human-scene interaction via prompted chain-of-contacts. arXiv preprint arXiv:2309.07918 (2023)
53. Xie, Y., Jampani, V., Zhong, L., Sun, D., Jiang, H.: Omnicontrol: Control any joint at any time for human motion generation. In: The Twelfth International Conference on Learning Representations (2024)
54. Xu, P., Shang, X., Zordan, V., Karamouzas, I.: Composite motion learning with task control. ACM Transactions on Graphics **42**(4) (2023). https://doi.org/10.1145/3592447
55. Xu, P., Xie, K., Andrews, S., Kry, P.G., Neff, M., McGuire, M., Karamouzas, I., Zordan, V.: AdaptNet: Policy adaptation for physics-based character control. ACM Transactions on Graphics **42**(6) (2023). https://doi.org/10.1145/3618375
56. Yang, H., Su, K., Zhang, Y., Chen, J., Qian, K., Liu, G., Gan, C.: Unimumo: Unified text, music, and motion generation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 25615–25623 (2025)
57. Yao, H., Song, Z., Chen, B., Liu, L.: Controlvae: Model-based learning of generative controllers for physics-based characters. ACM Transactions on Graphics (TOG) **41**(6), 1–16 (2022)
58. Yao, H., Song, Z., Zhou, Y., Ao, T., Chen, B., Liu, L.: Moconvq: Unified physics-based motion control via scalable discrete representations. ACM Transactions on Graphics (TOG) **43**(4), 1–21 (2024)
59. Yuan, Y., Song, J., Iqbal, U., Vahdat, A., Kautz, J.: Physdiff: Physics-guided human motion diffusion model. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 16010–16021 (2023)
60. Zhang, M., Cai, Z., Pan, L., Hong, F., Guo, X., Yang, L., Liu, Z.: Motiondiffuse: Text-driven human motion generation with diffusion model. IEEE transactions on pattern analysis and machine intelligence **46**(6), 4115–4128 (2024)
61. Zhang, Y., Tzeng, E., Du, Y., Kislyuk, D.: Large-scale reinforcement learning for diffusion models. In: European Conference on Computer Vision. pp. 1–17. Springer (2024)
62. Zhao, K., Li, G., Tang, S.: DartControl: A diffusion-based autoregressive motion model for real-time text-driven motion control. In: The Thirteenth International Conference on Learning Representations (ICLR) (2025)