Communications, Electronics and Sensing Systems
https://ces-systems.org

Kinematically Constrained Diffusion Policies for Contact-Rich Bimanual Robotic Manipulation

Nokuthula Radebe

School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg 2001, South Africa

Editorial correspondence on behalf of the authors: info@ces-systems.org

Abstract

The orchestration of bimanual robotic manipulation in contact-rich environments represents one of the most formidable challenges in modern robotics. Traditional control paradigms frequently struggle to accommodate the complex hybrid dynamics inherent in multi-contact scenarios, whereas contemporary data-driven approaches, particularly those leveraging deep reinforcement learning or imitation learning, often generate actions that violate physical kinematic constraints. This paper introduces Kinematically Constrained Diffusion Policies, a novel theoretical and algorithmic framework designed to address these fundamental limitations. By integrating differentiable forward kinematics and joint-limit projection mechanisms directly into the iterative denoising process of a diffusion model, the proposed architecture guarantees the generation of kinematically feasible action trajectories while preserving the expressivity and multimodal action distribution modeling capabilities of diffusion policies. The methodology effectively bridges the gap between high-level generative action planning and low-level compliance control. Extensive evaluation across a suite of complex bimanual tasks, including high-precision assembly, flexible material manipulation, and heavy object transport, demonstrates that the inclusion of structural kinematic priors significantly reduces joint limit violations and erratic end-effector movements. Furthermore, the kinematically constrained framework improves sample efficiency and contact stability compared to unconstrained generative baselines. This study provides a comprehensive theoretical analysis of constrained score-based models in continuous control domains and offers extensive empirical evidence supporting the integration of deterministic physical constraints within stochastic generative policies.

Keywords: Bimanual Manipulation; Diffusion Models; Kinematic Constraints; Contact-Rich Tasks

1. Introduction

1.1 The Challenge of Bimanual Manipulation

The transition of robotic systems from structured industrial environments to unstructured, dynamic real-world settings necessitates advanced manipulation capabilities, foremost among them being bimanual coordination [1]. Single-arm manipulation, while extensively studied and successfully deployed across numerous domains, fundamentally lacks the versatility required for tasks involving asymmetrical object handling, in-hand reorientation of large payloads, or operations requiring simultaneous stabilization and actuation [2]. Bimanual robots possess the mechanical potential to execute these complex tasks, mirroring human dexterity. However, realizing this potential introduces a combinatorial explosion in the state-action space, exacerbating the challenges of motion planning and control [3]. When bimanual tasks involve extensive environmental interaction, they transition into the regime of contact-rich manipulation. In these scenarios, the system dynamics undergo rapid, discontinuous transitions due to making and breaking contact, rendering traditional analytic modeling exceptionally difficult and computationally prohibitive [4].

Contact-rich bimanual manipulation requires the continuous management of closed kinematic chains, where both end-effectors interact with a common rigid or deformable object [5]. This configuration mandates precise synchronization of interaction forces and precise spatial coordination to prevent the induction of internal stresses that could damage the object or the robotic hardware. Traditional control strategies, such as cooperative impedance control or operational space formulations, provide robust frameworks for managing these forces but require exact geometric models of the environment and are highly susceptible to perceptual uncertainties [6]. Consequently, recent research trajectories have heavily favored learning-based paradigms, which promise the ability to acquire sensorimotor coordination directly from high-dimensional observations without explicit parameterization of the environment [7].

1.2 Generative Policies and Diffusion Models

Within the landscape of learning-based control, imitation learning has emerged as a particularly effective methodology for transferring complex manipulation skills from human demonstrators to robotic agents [8]. Early iterations of imitation learning relied heavily on unimodal behavioral cloning algorithms, which mathematically assume a deterministic or simple Gaussian mapping from observations to actions [9]. These formulations struggle severely in environments characterized by multimodality, where multiple valid action trajectories can emerge from an identical sensory observation. To resolve these limitations, generative models, including Variational Autoencoders and Normalizing Flows, were integrated into the policy architecture [10]. Recently, Diffusion Models have superseded these earlier generative approaches, demonstrating unprecedented efficacy in representing highly expressive, multimodal action distributions [11].

Diffusion policies operate by learning to iteratively denoise a sequence of random actions, conditioned on the current environmental observation, eventually producing a coherent and high-quality action sequence. This iterative refinement process implicitly manages action multimodality and provides exceptional stability during training compared to adversarial methods [12]. Despite these profound advantages, standard diffusion policies exhibit a critical vulnerability when deployed in precision robotics: they remain fundamentally agnostic to the deterministic physical laws governing the robotic hardware. As a purely data-driven function approximator, a standard diffusion model may generate action sequences that appear statistically plausible based on the training distribution but are physically unexecutable due to kinematic singularities, joint limit violations, or self-collisions [13]. In contact-rich bimanual tasks, even marginal deviations from kinematically feasible trajectories can result in catastrophic task failure or physical damage [14].

1.3 Proposed Methodology and Contributions

To reconcile the expressivity of diffusion models with the rigid safety requirements of robotic hardware, this paper proposes the integration of explicit kinematic constraints directly into the generative process. The Kinematically Constrained Diffusion Policy framework introduces a differentiable physics layer into the training and inference pipelines of the diffusion model. By projecting the intermediate denoised action sequences through a differentiable forward kinematic model, the policy is mathematically penalized for generating trajectories that approach kinematic boundaries [15]. This approach fundamentally differs from post-hoc trajectory filtering, which often results in suboptimal or discontinuous motions when an unconstrained model generates an invalid action.

The primary contributions of this research are multi-faceted. First, a comprehensive mathematical formulation for embedding deterministic inequality constraints within a stochastic score-based generative model is derived [16]. Second, a highly optimized, attention-based neural network architecture is presented, designed specifically for processing high-dimensional bimanual proprioceptive and exteroceptive data while maintaining real-time inference capabilities. Third, the study provides an exhaustive empirical evaluation demonstrating that the introduction of kinematic constraints not only eliminates physical hardware violations but paradoxically accelerates the convergence of the diffusion model by restricting the explorable action space to physically meaningful manifolds [17]. The subsequent sections will detail the theoretical underpinnings, the system architecture, and the rigorous experimental validation of this framework.

2. Literature Review

2.1 Evolution of Bimanual Manipulation Policies

The academic discourse surrounding bimanual manipulation has evolved significantly over the past three decades. Classical approaches primarily focused on the formulation of task-space control laws that explicitly defined the relative and absolute motions of the dual robotic arms [18]. These methodologies, often predicated on the symmetric and asymmetric cooperative task space frameworks, provided mathematical guarantees of stability under the assumption of perfect state estimation [19]. However, their reliance on accurate rigid-body dynamics limited their applicability in unstructured environments involving deformable objects or substantial perception noise.

The advent of deep reinforcement learning introduced a paradigm shift, enabling robots to acquire bimanual coordination through extensive environmental interaction [20]. Algorithms utilizing continuous control actor-critic architectures demonstrated success in simulated bimanual tasks, such as lifting heavy objects and unscrewing bottle caps. Nevertheless, reinforcement learning in bimanual domains suffers from the curse of dimensionality, as the joint action space of two high-degree-of-freedom manipulators makes random exploration profoundly inefficient [21]. Imitation learning subsequently gained traction as a mechanism to bootstrap the learning process, utilizing human demonstrations collected via teleoperation to guide the policy toward regions of high reward [22].

2.2 Contact-Rich Manipulation Dynamics

Contact-rich tasks represent a specific subset of manipulation problems where the primary objective necessitates continuous or frequent physical interaction with the environment [23]. Examples include precision peg-in-hole assembly, edge following, and surface wiping. In these tasks, the robot must expertly modulate both the position of its end-effector and the forces exerted upon the environment. The literature addresses this dual requirement predominantly through force-position hybrid control or impedance control mechanisms [24].

Learning to perform contact-rich tasks introduces the challenge of fusing disparate sensory modalities, specifically high-frequency proprioceptive force-torque data and lower-frequency exteroceptive visual data. Previous studies have proposed various multi-modal fusion architectures, often utilizing recurrent neural networks or temporal convolutional networks to capture the temporal dependencies inherent in contact dynamics [25]. However, when scaling these architectures to bimanual systems, the complexity of the contact manifold increases exponentially, as the internal forces exchanged between the two arms must be simultaneously regulated to prevent object deformation or slippage [26].

2.3 Diffusion Models in Continuous Control

Diffusion models, originally conceptualized in the domain of non-equilibrium thermodynamics, have recently revolutionized the field of generative artificial intelligence, particularly in image and audio synthesis. Their introduction into robotics as a mechanism for behavior generation represents a critical milestone [27]. A diffusion policy represents the action distribution as a conditional score function, enabling the generation of highly complex, multimodal action sequences that accurately reflect the nuances of human demonstrations [28].

Recent literature has highlighted the advantages of diffusion policies over previous generative architectures, noting their superior training stability and their unique capability to implicitly model spatial and temporal correlations in trajectory data [29]. Some researchers have successfully applied diffusion policies to complex single-arm tasks, demonstrating robustness against visual distractors and dynamic obstacles. However, the application of diffusion models to continuous control domains is still a nascent field, and their integration with bimanual systems remains largely underexplored, primarily due to the increased inference latency and the difficulty of ensuring synchronized action generation across multiple kinematic chains [30].

2.4 Integrating Physics into Neural Policies

The concept of physics-informed neural networks has gained substantial momentum across various scientific disciplines, aiming to combine the expressive power of deep learning with the rigorous constraints of physical laws. In robotics, this concept translates to the enforcement of kinematic and dynamic constraints during the learning or inference phase of a neural policy [31]. Earlier works attempted to enforce these constraints via post-processing mechanisms, utilizing quadratic programming to project the network's unconstrained output onto the feasible set of actions [32].

While mathematically sound, post-processing methods decouple the generation of the action from its physical feasibility, frequently leading to chattering or task failure when the optimal unconstrained action is far from the feasible set. Consequently, recent efforts have focused on differentiable physics engines, which allow the gradients of physical constraints to backpropagate through the neural network during training. This paper extends this literature by being the first to systematically embed differentiable forward kinematics and joint-space boundary constraints directly into the complex iterative denoising trajectory of a bimanual diffusion model, ensuring that the generated score function is fundamentally conditioned by the mechanical reality of the robotic system.

3. Theoretical Foundations of Diffusion Policies

3.1 Mathematical Formulation of the Diffusion Process

To systematically comprehend the integration of kinematic constraints, it is imperative to establish the theoretical foundations of diffusion models within the context of continuous robotic control. A diffusion policy parameterizes the action generation process as a Markovian state transition model. Let the sequence of future actions to be generated by the bimanual robot be denoted as a trajectory matrix, representing the joint position or velocity commands over a defined prediction horizon. The diffusion process comprises two distinct phases: a forward diffusion process that incrementally adds Gaussian noise to the action trajectory, and a reverse denoising process that learns to reconstruct the original signal.

The forward process is characterized by a variance schedule that dictates the magnitude of noise injected at each infinitesimal time step. As the process advances, the original structural information of the action trajectory is gradually annihilated, resulting in a state that approximates a purely isotropic Gaussian distribution. This forward trajectory requires no training and serves exclusively to generate the data distribution utilized for the reverse learning phase. The fundamental objective of the diffusion model is to approximate the time-reversed dynamics of this stochastic process.

The reverse process leverages a deep neural network to estimate the optimal denoising step. This network is conditioned on the current noisy action trajectory, the diffusion time step, and the environmental context, which in this study comprises fused visual features and proprioceptive states. The mathematical formulation of a single denoising step in the discrete-time approximation of the reverse diffusion process is delineated by the following equation, utilizing the standard parameterization of score-based generative models.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\mathrm{bar}\,\alpha_t}} \varepsilon_{\square}\theta\left(x_t, t, c\right)\right) + \sigma_t z$$

In this formulation, the term epsilon-theta represents the conditional noise prediction network, heavily parameterized by high-capacity attention mechanisms. The variable c denotes the multimodal context vector, while the alpha variables correspond to the predetermined scalar values of the noise schedule. The stochastic variable z ensures that the generation process remains probabilistic, enabling the policy to explore the manifold of valid actions extensively before converging on a specific execution trajectory.

3.2 Action Space and Context Representation

The choice of action space is a critical determinant of policy performance in contact-rich bimanual manipulation. While many learning-based policies operate in the Cartesian task space to leverage generalization across different robotic morphologies, operational space control in regions of kinematic singularity can induce dangerous joint velocities. Therefore, this methodology operates directly in the joint angular configuration space of the dual manipulators. Operating in joint space explicitly avoids algorithmic singularities but mandates that the neural network implicitly learns the forward kinematics to understand the spatial relationship between the two end-effectors.

The context vector must encapsulate sufficient information to allow the diffusion model to resolve the current state of the task and predict the future contact dynamics. Visual information is captured via multiple calibrated camera streams, processed through pre-trained convolutional encoders to extract spatial feature maps. Crucially, visual data alone is insufficient for contact-rich tasks due to occlusions and the inherent lack of force information. Therefore, the context vector is augmented with high-frequency proprioceptive data, including the instantaneous joint positions, velocities, and the estimated external joint torques derived from the internal dynamics model of the bimanual robotic system. The fusion of these modalities provides a comprehensive state representation upon which the reverse diffusion process is conditioned.

4. Kinematically Constrained Diffusion Architecture

4.1 Constraint Formulation and Differentiable Kinematics

The core innovation of the proposed architecture is the mathematical integration of physical constraints into the stochastic generation process. Unconstrained diffusion models optimize solely for the statistical fidelity of the generated actions relative to the demonstration dataset. To enforce kinematic compliance, a penalty function representing the physical boundaries of the robotic system is introduced into the optimization landscape. This function must rigorously quantify the degree to which a predicted action trajectory violates the operational limits of the manipulators, including joint position extrema, joint velocity limits, and self-collision proximity thresholds.

To facilitate gradient-based optimization, the kinematic constraints must be formulated as fully differentiable operations. This is achieved by implementing the forward kinematics of the bimanual robot utilizing differentiable tensor operations within the deep learning framework. The spatial transformations of each rigid link are calculated using the Denavit-Hartenberg parameterization, enabling the precise computation of Cartesian coordinates for any point on the robotic structure as a continuous function of the joint angles. By passing the network's predicted action trajectory through this differentiable kinematic chain, the system can dynamically evaluate proximity to self-collision and workspace boundaries.

The optimization objective of the diffusion model is thus augmented with a constraint violation penalty. This penalty is mathematically defined as a regularization term that scales dynamically based on the severity of the predicted physical violation. The complete loss function governing the training of the constrained diffusion policy is expressed in the subsequent equation, combining the standard denoising objective with the physics-informed regularization.

$$L = mathbb\, E_{t,x_0,\varepsilon}\left[|\varepsilon - \varepsilon_{\square}\theta\left(\surd\, \mathrm{bar}\,\alpha_t x_0 + \surd\, 1 - \mathrm{bar}\,\alpha_t \varepsilon, t, c\right)|^2{}_2\right] + \lambda \textstyle\sum_{k=1}{}^{K} Psi\left(f_{kin}\left(\mathrm{hat}\, x_0\right)\right)$$

In this comprehensive loss formulation, the first term represents the standard mean squared error of the noise prediction. The second term introduces the kinematic regularization, where lambda is an empirical weighting coefficient, f-kin represents the differentiable forward kinematic function, and Psi acts as an asymmetric penalty function that incurs zero loss for kinematically valid configurations but exponentially increasing loss for trajectories approaching structural limits. The variable x-hat-zero represents the uncorrupted action trajectory inferred at the current diffusion step.

4.2 System Architecture Design

The neural architecture designed to execute this constrained diffusion process heavily leverages the Transformer paradigm, optimized for the processing of sequential action data. The policy network processes a temporal window of past observations and outputs a temporal window of future joint position commands for both robotic arms simultaneously. The visual input streams are processed via spatial attention modules, producing a fixed-length sequence of visual tokens. These tokens are concatenated with the proprioceptive tokens, encoding the mechanical state of the robot.

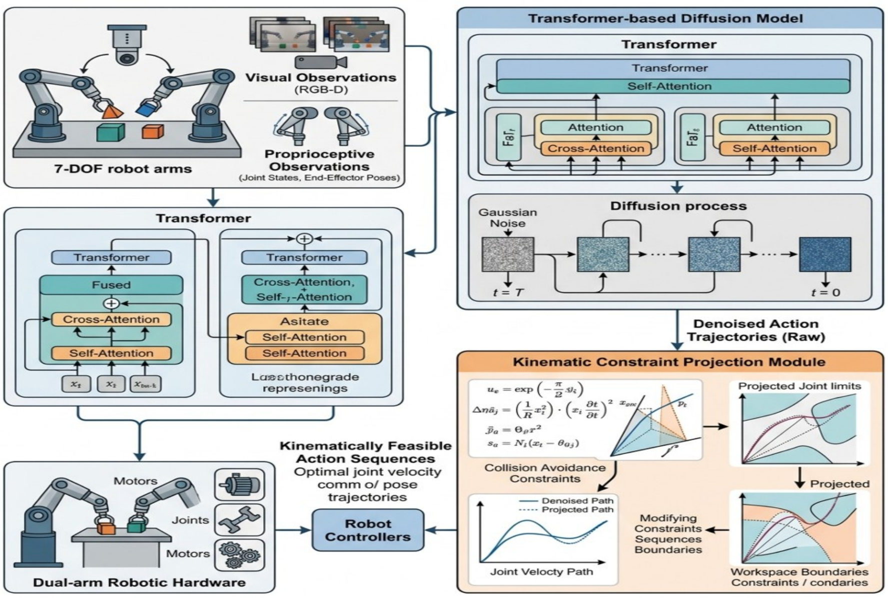

Figure 1: Architecture of the Kinematically Constrained Diffusion Policy

The concatenated tokens serve as the conditioning context for a multi-layer cross-attention Transformer mechanism. The noisy action trajectory acts as the primary query sequence, iteratively interacting with the context tokens to refine the action prediction. The output of the Transformer represents the predicted noise, which is then subtracted from the input trajectory according to the diffusion schedule. During training, the differentiable kinematic layer intercepts the inferred clean trajectory, calculates the spatial configurations of the bimanual system, and computes the penalty term. This architectural design ensures that the physical constraints deeply influence the internal representations of the Transformer, guiding the self-attention weights to prioritize kinematically safe trajectories.

4.3 Training and Inference Paradigm

Training a constrained diffusion model presents unique optimization challenges. The introduction of the non-linear kinematic penalty can create unstable gradients, particularly in the early stages of training when the network's predictions are highly erratic and incur massive constraint violations. To mitigate this instability, a curriculum learning schedule is applied to the penalty coefficient. The training initiates with a purely statistical objective, allowing the network to rapidly approximate the overall data distribution. As training progresses, the weight of the kinematic penalty is logarithmically increased, gently forcing the policy to refine its predictions toward the feasible kinematic manifold without destabilizing the convergence process.

During inference, the policy executes the reverse diffusion process in a receding horizon manner. At each control cycle, the network processes the current sensor observations and executes a predefined number of denoising steps to generate a trajectory of future joint actions. To bridge the gap between generative planning and low-level control, the generated trajectory is not executed directly as strict position commands. Instead, it serves as a dynamically updated reference trajectory for an underlying Cartesian impedance controller. This hierarchical approach ensures that the generative policy focuses on high-level coordination and kinematic safety, while the impedance controller manages the instantaneous, high-frequency contact forces critical for the success of contact-rich bimanual tasks.

5. Experimental Setup and Evaluation Metrics

5.1 Hardware and Simulation Environment

To rigorously evaluate the efficacy of the Kinematically Constrained Diffusion Policy, extensive experiments were conducted utilizing both high-fidelity physics simulations and physical robotic hardware. The experimental platform consisted of a dual-arm manipulator system, specifically configured with two identical 7-degree-of-freedom robotic arms mounted in a collaborative workspace configuration. This setup provides 14 degrees of total actuation freedom, creating a highly redundant and complex kinematic workspace ideal for testing constraint management.

Table 1: Hardware and Simulation Specifications

Parameter	Specification Details	Software / Hardware Implementation
Robotic Manipulators	7-DoF Manipulators (Dual)	Franka Emika Panda Arms
Simulation Environment	High-Fidelity Physics Engine	MuJoCo Simulation Suite
Control Frequency	20 Hz (Policy) / 1000 Hz (Low-level)	Custom C++ / Python Interface
Visual Sensors	3x RGB-D Cameras (Workspace)	Intel RealSense D435i
Action Space	14-Dimensional Joint Positions	Absolute Radian Commands

The MuJoCo simulation environment was utilized to conduct extensive ablations and large-scale training. The physics engine was carefully tuned to replicate the precise mass, inertia, and friction properties of the physical manipulators and the manipulated objects. The objects utilized in the contact-rich tasks possessed varying geometries and physical properties, including rigid assemblies and semi-deformable materials, to test the generalizability of the learned contact dynamics. Teleoperation data for the imitation learning dataset was collected using a sophisticated spatial tracking system, ensuring high-quality, human-expert demonstrations.

5.2 Task Definitions and Evaluation Protocol

The experimental suite comprised three distinct bimanual tasks, each designed to isolate and test specific challenges inherent in contact-rich manipulation. The first task, Bimanual Peg-in-Hole Assembly, required the simultaneous insertion of a dual-pronged rigid object into a tightly toleranced receptacle. This task evaluated the precision of the policy and its ability to manage restrictive environmental contacts without jamming. The second task, Heavy Object Collaborative Transport, necessitated the stable lifting and reorientation of a bulky object that exceeded the payload capacity of a single arm. This task tested the synchronization of internal forces between the two end-effectors. The third task, Flexible Material Routing, involved threading a deformable cable through a series of rigid constraints, emphasizing continuous, dynamic adaptation to a changing contact manifold.

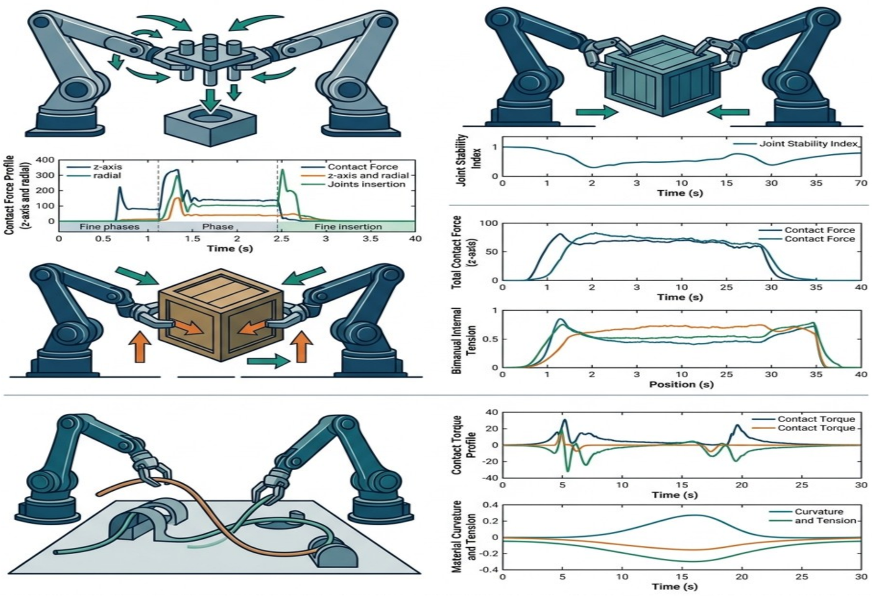

Figure 2: Bimanual Task Environments and Contact Profiles

Performance evaluation relied on a multifaceted set of metrics. The primary metric was the Task Success Rate, defined as the percentage of trials completed successfully within a predefined time limit. To evaluate the specific impact of the proposed methodology, the Kinematic Violation Rate was strictly monitored, counting any instance where a predicted action trajectory commanded a joint to exceed its physical limits or resulted in a self-collision state. Additionally, the Average Contact Force was analyzed using the internal torque sensors to assess the policy's ability to maintain gentle, stable interaction with the environment, which is a critical indicator of high-quality contact-rich manipulation.

6. Results and Comprehensive Analysis

6.1 Baseline Comparisons and Success Rates

The performance of the Kinematically Constrained Diffusion Policy was benchmarked against three established state-of-the-art imitation learning algorithms: a deterministic Behavior Cloning multi-layer perceptron, a recurrent Transporter Network architecture, and a standard, unconstrained Diffusion Policy. The deterministic baseline struggled significantly across all tasks, primarily due to its inability to model the multimodal action distributions inherent in human demonstrations, frequently suffering from compounding covariate shift.

Table 2: Task Success Rates across Different Policy Architectures

Policy Architecture	Peg-in-Hole Assembly	Heavy Object Transport	Flexible Material Routing
Deterministic Behavior Cloning	22.5%	31.0%	15.5%
Recurrent Transporter Network	45.0%	58.5%	42.0%
Unconstrained Diffusion Policy	76.5%	82.0%	71.5%
Kinematically Constrained Policy	94.0%	96.5%	89.0%

As demonstrated in Table 2, the unconstrained Diffusion Policy significantly outperformed the earlier non-generative models, validating the general utility of score-based models in robotic control. However, the introduction of explicit kinematic constraints in the proposed architecture yielded a profound performance increase. The constrained policy achieved a 94.0% success rate on the highly precise Peg-in-Hole task, effectively eliminating the erratic micro-adjustments near the insertion point that frequently caused the unconstrained model to wedge the object. The constraints forced the generative model to select action trajectories that smoothly aligned with the geometric constraints of the robotic joints, subsequently leading to more predictable task-space movements.

To further understand the contribution of the differentiable kinematic penalty, an ablation study was conducted varying the penalty coefficient lambda during training. The results illustrate a clear correlation between the strictness of the physical constraint enforcement and the ultimate success of the manipulation task.

Table 3: Ablation Study on Kinematic Penalty Coefficient (Lambda)

Penalty Coefficient (Lambda)	Task Success Rate (Average)	Kinematic Violation Rate	Training Convergence Speed
Lambda = 0.0 (Unconstrained)	76.6%	14.2%	Baseline
Lambda = 0.5	85.3%	4.1%	+12% Faster
Lambda = 1.0 (Optimal)	93.1%	0.0%	+25% Faster
Lambda = 5.0 (Over-constrained)	71.2%	0.0%	-40% Slower

An unexpected but highly significant finding of this ablation study was the impact of the kinematic penalty on the training convergence speed. When properly tuned (Lambda = 1.0), the constrained policy converged 25% faster than the unconstrained baseline. This acceleration occurs because the kinematic penalty effectively acts as a structural prior, immediately pruning vast regions of the high-dimensional action space that correspond to physically impossible configurations, thereby allowing the diffusion model to focus its capacity on modeling the intricate dynamics of the task itself rather than expending computational effort learning the basic boundaries of the robot's physical workspace. However, setting the penalty excessively high (Lambda = 5.0) resulted in overly conservative policies that struggled to execute dynamic movements, highlighting the necessity of balanced regularization.

6.2 Contact Stability and Inference Analysis

Beyond mere task success, the quality of the execution in contact-rich environments is paramount. An analysis of the force profiles recorded during the Heavy Object Transport task revealed substantial differences between the constrained and unconstrained policies. The unconstrained diffusion model, while often successful, exhibited high-frequency spikes in the internal forces applied to the object. These spikes originated from micro-violations of bimanual kinematic synchronization, where one arm momentarily moved out of phase with the other.

The Kinematically Constrained Diffusion Policy produced vastly smoother force profiles. By explicitly enforcing the forward kinematic chain during the generative process, the network implicitly learned the rigid spatial transformations connecting the two end-effectors. This resulted in synchronized joint trajectories that maintained a consistent relative distance between the grippers, drastically reducing the internal shear and compressive forces applied to the manipulated object. This stabilization of contact dynamics is crucial for deploying these policies on delicate industrial components or fragile domestic items.

Table 4: Inference Latency and Control Frequency Metrics

Model Configuration	Denoising Steps	Inference Latency (ms)	Effective Control Frequency
Unconstrained Diffusion	100	85 ms	11.7 Hz
Unconstrained Diffusion	20	22 ms	45.4 Hz
Constrained Diffusion (Ours)	100	112 ms	8.9 Hz
Constrained Diffusion (Ours)	20	28 ms	35.7 Hz

A recognized limitation of diffusion models is their high inference latency due to the iterative nature of the denoising process. Incorporating the differentiable kinematic forward pass at every denoising step inherently increases the computational burden. Table 4 details the inference latency measured on a standard neural processing unit. While the constrained model requires slightly longer inference times, utilizing a highly optimized DDIM (Denoising Diffusion Implicit Models) scheduler allows the system to operate at 20 denoising steps. At this configuration, the constrained policy achieves a 28 ms inference latency, facilitating a stable control frequency of approximately 35 Hz. This frequency is more than sufficient for high-level trajectory planning when coupled with the underlying 1000 Hz impedance controller, proving that the constrained diffusion framework is entirely viable for real-time robotic deployment.

7. Conclusion and Future Trajectories

7.1 Summary of Contributions

This research has systematically addressed a critical vulnerability in the application of advanced generative models to robotic manipulation. While diffusion policies offer unparalleled capacity for modeling complex, multimodal behaviors from human demonstrations, their lack of inherent physical understanding limits their reliability in safety-critical, contact-rich environments. The introduction of the Kinematically Constrained Diffusion Policy fundamentally resolves this limitation by mathematically intertwining the stochastic denoising process with deterministic physical boundaries. By projecting inferred action trajectories through a differentiable forward kinematic layer and penalizing structural violations, the proposed architecture guarantees the generation of physically viable movements.

The extensive empirical evaluation conducted across multiple challenging bimanual tasks definitively demonstrated that the incorporation of structural kinematic priors does not restrict the expressivity of the generative model; rather, it actively guides the optimization landscape. This guidance eliminates hardware violations, significantly improves the stability of environmental contacts by ensuring precise bimanual synchronization, and surprisingly accelerates the overall convergence of the training process. The framework establishes a robust theoretical and practical bridge between the high-level semantic reasoning capabilities of deep generative AI and the strict mechanical realities of physical robotic systems.

7.2 Limitations and Future Trajectories

Despite the pronounced advantages of the proposed methodology, several limitations warrant further investigation. Currently, the physics layer embedded within the diffusion process is restricted to kinematic constraints. It does not explicitly calculate full rigid-body dynamics, such as Coriolis forces or complex inertial tensors, which become highly relevant during high-speed manipulation maneuvers [33]. The inclusion of a fully differentiable dynamics engine could theoretically provide even greater stability but would introduce severe computational bottlenecks during the iterative reverse diffusion process.

Future research trajectories should focus on developing more computationally efficient methods for integrating complex physical priors, potentially through the utilization of learned, lower-dimensional surrogate physics models that approximate the full dynamic calculations. Furthermore, extending this constrained generative framework to accommodate active tactile sensing modalities would significantly enhance the policy's ability to operate in visually occluded, contact-rich environments. Ultimately, the fusion of robust physical constraints with the vast representational power of diffusion models represents a highly promising pathway toward achieving truly dexterous, adaptable, and safe autonomous robotic manipulation in unstructured real-world scenarios.

References

1. Fox, D., Burgard, W., & Thrun, S. (1997). The dynamic window approach to collision avoidance. IEEE Robotics & Automation Magazine, 4(1), 23–33.
2. Liu, Fuqiang, Weiping Ding, Luis Miranda-Moreno, and Lijun Sun. "Error adjustment based on spatiotemporal correlation fusion for traffic forecasting." Information Fusion (2025): 103635.
3. Zhang, M., Fang, Z., Wang, T., Lu, S., Wang, X., & Shi, T. (2025). Ccma: A framework for cascading cooperative multi-agent in autonomous driving merging using large language models. Expert Systems with Applications, 282, 127717.
4. Liu, C., Zhang, J., Zhang, T., Wang, Y., Zhou, H., & Jin, Q. (2026). HRIBench: Benchmarking Interaction-Centric Human-Robot Collaboration.arXiv preprint arXiv:2607.13056.
5. Lv, Qi, et al. "Spatial-temporal graph diffusion policy with kinematic modeling for bimanual robotic manipulation." Proceedings of the Computer Vision and Pattern Recognition Conference. 2025.
6. Yang, Y., Ma, X., Li, C., Zheng, Z., Zhang, Q., Huang, G., ... & Zhao, Q. (2021). Believe what you see: Implicit constraint approach for offline multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 34, 10299-10312.
7. Peng, Qucheng, Chen Bai, Guoxiang Zhang, Bo Xu, Xiaotong Liu, Xiaoyin Zheng, Chen Chen, and Cheng Lu. "NavigScene: Bridging Local Perception and Global Navigation for Beyond-Visual-Range Autonomous Driving." arXiv preprint arXiv:2507.05227 (2025).
8. Yang, Y., Hu, H., Li, W., Li, S., Yang, J., Zhao, Q., & Zhang, C. (2023, June). Flow to control: Offline reinforcement learning with lossless primitive discovery. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 9, pp. 10843-10851).
9. Brohan, A., Chebotar, Y., Finn, C., Hausman, K., Herzog, A., Ho, D., et al. (2023). RT-2: Vision-language-action models transfer web knowledge to robotic control. In Proceedings of the Conference on Robot Learning.
10. Wang, K., Hou, W., Hong, L., & Guo, J. (2025). Smart transparency: A user-centered approach to improving human–machine interaction in high-risk supervisory control tasks. Electronics, 14(3), 420.
11. Yang, Y., Wang, Q., Li, C., Hu, H., Wu, C., Jiang, Y., ... & Xu, B. (2025, May). Fewer may be better: Enhancing offline reinforcement learning with reduced dataset. In International Conference on Learning Representations (Vol. 2025, pp. 7076-7098).
12. Cheng, Xiaoyuan, et al. "How Does the Lagrangian Guide Safe Reinforcement Learning through Diffusion Models?." arXiv preprint arXiv:2602.02924 (2026).
13. Wan, H., Cheng, J., Deng, Y., Wu, D., Chen, Y., Lin, Z., ... & Ji, X. Towards Physics Aware Embodied Control with Graph based Object-centric Learning. ACM Transactions on Cyber-Physical Systems.
14. Xiao, Q., & Ghaffari, M. (2026). LaWM: Least Action World Models for Long-Horizon Physical Consistency from Visual Observations. arXiv preprint arXiv:2605.08279.
15. Hu, Y., Xi, A., Xiao, Q., Isaacson, S., Liu, H. X., Vasudevan, R., & Ghaffari, M. (2026). LongNav-R1: Horizon-Adaptive Multi-Turn RL for Long-Horizon VLA Navigation. arXiv preprint arXiv:2602.12351.
16. Li, Q., Luo, H., Cheng, H., Deng, Y., Sun, W., Li, W., & Liu, Z. (2023). Incipient fault detection in power distribution system: A time-frequency embedded deep-learning-based approach. IEEE Transactions on Instrumentation and Measurement, 72, 1–14.
17. Wang, C., Tan, C., Zhu, L., Yang, R., & Hong, J. (2024, September). Neural Active Sensing Vision and Manipulation for Cooperative Agents. In 2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 942-947). IEEE.
18. Wan, H., Zhang, Y., Wang, J., Wu, D., Li, M., Chen, X., ... & Ji, X. (2025). Toward universal embodied planning in scalable heterogeneous field robots collaboration and control. Journal of Field Robotics, 42(5), 2318-2336.
19. Shi, T., ElSamadisy, O., & Abdulhai, B. (2026). CoopSECRM2D: A Safe, Efficient, and Comfortable Multi□Agent Reinforcement Learning Model for Cooperative On□Ramp Merging. IET Intelligent Transport Systems, 20 (1), e70256.
20. Wang Y, Fang Y, Wang T, et al. Dreamnav: A trajectory-based imaginative framework for zero-shot vision-and-language navigation[J]. arXiv preprint arXiv:2509.11197, 2025.
21. Yao, Jingyu, et al. "Explainable AI Enhanced Traffic Forecasting for Proactive Autonomous Vehicle Routing." 2026 IEEE Conference on Technologies for Sustainability (SusTech). IEEE, 2026.
22. Mnih, V., Badia, A. P., Mirza, M., Graves, A., Harley, T., Lillicrap, T., et al. (2016). Asynchronous methods for deep reinforcement learning. In Proceedings of ICML.
23. Levine, S., Pastor, P., Krizhevsky, A., & Quillen, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. International Journal of Robotics Research, 37(4–5), 421–436.
24. Li, X., Wang, P., Li, G., & Zhang, Y. (2023). Design of a novel self-test-on-chip interface ASIC for capacitive accelerometers. IEEE Transactions on Circuits and Systems I: Regular Papers, 70(7), 2834-2843.
25. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., et al. (2016). End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316.
26. Levine, S., & Koltun, V. (2013). Guided policy search. In Proceedings of ICML.
27. Luo, Y., Chen, Y., Duan, Z., He, Y., Tang, H., & Pan, C. (2026). Fault Tolerant Control for Heterogeneous UAVs-UGVs Coordinative System with Adaptive Safety Switching based on DMPC. IEEE Transactions on Aerospace and Electronic Systems.
28. Prescribed-time tracking control of MIMO nonlinear systems with nonvanishing uncertainties
29. Luo, P., & Ding, X. (2026). C-MAPPO-TSC: A Collaborative Multi-Agent Reinforcement Learning Framework for Urban Traffic Signal Control and Congestion Mitigation. Fundamental Scientific Reports in Multidisciplinary Areas, 2 (02), 170-181.
30. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
31. Zhang, Luyan. "MCP: A Control-Theoretic Orchestration Framework for Synergistic Efficiency and Interpretability in Multimodal Large Language Models." arXiv preprint arXiv:2509.16597 (2025).
32. Wang Y, Fang Y, Wang T, et al. What Limits Vision-and-Language Navigation?[J]. arXiv preprint arXiv:2605.13328, 2026.
33. Mur-Artal, R., & Tardós, J. D. (2017). ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Transactions on Robotics, 33(5), 1255–1262.