STRIDE: Structured Lagrangian and Stochastic Residual Dynamics via Flow Matching

Prakrut Kotecha[1], Ganga Nair B[1], Shishir Kolathaya[2]

Abstract— **Robotic systems operating in unstructured environments must operate under significant uncertainty arising from intermittent contacts, frictional variability, and unmodeled compliance. While recent model-free approaches have demonstrated impressive performance, many deployment settings still require predictive models that support planning, constraint handling, and online adaptation. Analytical rigid-body models provide strong physical structure but often fail to capture complex interaction effects, whereas purely data-driven models may violate physical consistency, exhibit data bias, and accumulate long-horizon drift. In this work, we propose STRIDE, a dynamics learning framework that explicitly separates conservative rigid-body mechanics from uncertain, effectively stochastic non-conservative interaction effects. The structured component is modeled using a Lagrangian Neural Network (LNN) to preserve energy-consistent inertial dynamics, while residual interaction forces are represented using Conditional Flow Matching (CFM) to capture multi-modal interaction phenomena. The two components are trained jointly end-to-end, enabling the model to retain physical structure while representing complex stochastic behavior. We evaluate STRIDE on systems of increasing complexity, including the Unitree Go1 quadruped, and the Unitree G1 humanoid. Results show 20% reduction in long-horizon prediction error and 30% reduction in contact force prediction error compared to deterministic residual baselines, supporting more reliable model-based control in uncertain robotic environments.**

I. Introduction

Robotic systems are increasingly expected to operate outside controlled laboratory settings and within unstructured, dynamic environments. In such conditions, robots must contend with significant uncertainty arising from intermittent contacts, frictional variability, unmodeled compliance, and actuator nonlinearities [1]. These effects are particularly pronounced in legged and humanoid systems, where small disturbances during contact can lead to large deviations in system behavior.

Model-based control frameworks offer planning capabilities unlike purely reactive policies like RL by explicitly leveraging system dynamics. When a sufficiently accurate model is available, control methods such as Model Predictive Control (MPC) and trajectory optimization can incorporate task objectives at runtime, enforce state and input constraints, and locally replan under distribution shift. In practice, however, obtaining dynamics models that are both physically consistent and expressive enough to capture real-world interaction uncertainty remains a central challenge.

[1]G. Nair, P. Kotecha, are with the Robert Bosch Center for Cyber-Physical Systems, Indian Institute of Science, Bengaluru.

[2]S. Kolathaya is with the Robert Bosch Center for Cyber-Physical Systems and the Department of Computer Science & Automation, Indian Institute of Science, Bengaluru.

Analytical rigid-body models provide strong physical structure grounded in first-principles [2]–[4]. However, deriving high-fidelity analytical models for complex robotic systems is often difficult and computationally expensive, and their accuracy deteriorates in the presence of frictional changes, impacts, and unmodeled compliance. Purely data-driven approaches have emerged as an alternative by learning system dynamics directly from trajectory data [5], [6]. While expressive, such models typically lack physical inductive bias, require large datasets for generalization, and may violate fundamental physical constraints, leading to energy inconsistency and compounding long-horizon prediction errors [7].

To bridge analytical and learning-based paradigms, physics-informed neural architectures embed structural priors directly into learned models. In particular, Lagrangian Neural Networks (LNNs) parameterize system dynamics through learned kinetic and potential energy functions, enforcing the Euler-Lagrange equations by construction [8], [9]. These approaches typically represent non-conservative effects such as friction and contact impulses using deterministic residual terms [10]–[12]. Such a purely deterministic representation of non-conservative forces fails to capture the complexity inherent in these interactions.

Recent efforts to capture stochasticity in prediction have explored fully generative transition models, including diffusion and score-based approaches [13]–[15] which often lack the structural decompositions required for planning, and introduces prohibitive computational overhead. Conditional Flow Matching (CFM) offers an attractive alternative: by enabling direct, efficient sampling from complex conditional distributions without the multi-step denoising [16].

Motivated by these observations, we propose **STRIDE**, a dynamics learning framework that models conservative rigid-body mechanics using an LNN to preserve the inductive biases of analytical mechanics, and stochastic non-conservative effects modeled using CFM to capture multi-modal variability efficiently. By doing so, STRIDE preserves physical consistency while reducing the modeling burden on the generative component.

II. Problem Formulation

A. Mechanical System Model

We consider a robotic system defined by generalized coordinates $\mathbf{q} \in \mathbb{R}^n$ and generalized velocities $\dot{\mathbf{q}} \in \mathbb{R}^n$. In an ideal, undisturbed environment, the system evolves according to classical Lagrangian mechanics:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial \mathcal{L}}{\partial \mathbf{q}} = \boldsymbol{\tau} + \mathbf{F}_{\text{ext}}, \tag{1}$$

where $\boldsymbol{\tau} \in \mathbb{R}^n$ represents the known control inputs and $\mathbf{F}_{\text{ext}}$ represents the sum of all external, non-conservative, and unmodeled forces. The scalar Lagrangian ($\mathcal{L}$) is defined as:

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) = \mathcal{T}(\mathbf{q}, \dot{\mathbf{q}}) - \mathcal{V}(\mathbf{q}), \tag{2}$$

where $\mathcal{T}$ and $\mathcal{V}$ denote the kinetic and potential energy, respectively.

By expanding (1), we can arrive at the standard rigid-body dynamics form:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) = \boldsymbol{\tau} + \mathbf{F}_{\text{ext}}, \tag{3}$$

where $\mathbf{M}(\mathbf{q})$ is the positive-definite mass matrix, $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ captures Coriolis and centrifugal effects, and $\mathbf{g}(\mathbf{q})$ represents gravitational forces.

Equation (3) motivates a useful modeling decomposition of robot dynamics. The rigid-body terms capture the dominant, geometry-dependent mechanics arising from the robot's morphology (e.g., mass distribution and inertial coupling), while the external force term $\mathbf{F}_{\text{ext}}$ aggregates non-conservative and environment-dependent effects. $\mathbf{F}_{\text{ext}}$ often exhibit multi-modal behavior near contact transitions. STRIDE takes advantage of this decomposition.

B. Learning Objective

Our goal is to learn a predictive model $\ddot{\mathbf{q}} = f_\theta(\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\tau})$ that does not sacrifice physical consistency for expressive power. We propose a structural decomposition of the predicted acceleration:

$$\ddot{\mathbf{q}}_{\text{pred}} = f_{\text{LNN}}(\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\tau}) + \mathbf{M}^{-1}(\mathbf{q})\boldsymbol{\epsilon}_{\text{CFM}}(\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\tau}), \tag{4}$$

where:

1) f_{LNN} is a deterministic **LNN** that learns the conservative prior by parameterizing $\mathbf{M}(\mathbf{q})$ and $\mathcal{V}(\mathbf{q})$.
2) $\boldsymbol{\epsilon}_{\text{CFM}}$ is a **CFM** residual that estimates the unstructured dynamics using a stochastic generative process.

Rather than learning the conservative and residual components independently, we train the full model jointly in an end-to-end manner under a unified objective.

III. Methodology

Lagrangian Prior: To ensure physical consistency, STRIDE constructs the mass matrix via a Cholesky factorization

$$\mathbf{M}_\theta(\mathbf{q}) = \mathbf{L}_\theta(\mathbf{q})\mathbf{L}_\theta(\mathbf{q})^\top,$$

where $\mathbf{L}_\theta$ is lower triangular and its diagonal entries are enforced to be positive using a softplus activation. The kinetic energy is then computed as

$$\mathcal{T}(\mathbf{q}, \dot{\mathbf{q}}) = \dot{\mathbf{q}}^\top \mathbf{M}_\theta(\mathbf{q})\dot{\mathbf{q}}.$$

We also learn a scalar-valued network to predict the potential energy $\mathcal{V}_\theta(\mathbf{q})$. Together with (2) and (1), this formulation encourages f_{LNN} to capture the conservative mechanics of the system, reducing the risk of energy drift commonly observed in purely data-driven models.

Generative Residuals via Flow Matching: While the LNN preserves conservative dynamics, deterministic modeling of $\mathbf{F}_{\text{ext}}$ introduces a failure mode in contact-rich regimes. Let $p^*(\mathbf{F}_{\text{ext}} \mid \mathbf{c})$, denote the true distribution of non conservative forces (where context $\mathbf{c} = (\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\tau})$), which is often multi-modal under intermittent contact, frictional changes etc. A common modeling strategy is to approximate the residual forces using a deterministic regressor trained with a mean-squared error objective. Under the standard L_2 risk,

$$\min_f \ \mathbb{E}\left[\|\mathbf{F}_{\text{ext}} - f(\mathbf{c})\|^2\right], \tag{5}$$

the optimal predictor is the conditional mean

$$f^*(\mathbf{c}) = \mathbb{E}_{p^*}[\mathbf{F}_{\text{ext}} \mid \mathbf{c}]. \tag{6}$$

This mean may lie between physically distinct modes. In robotic systems, such averaging can smooth over discontinuous interaction phenomena, for example, when a foot either slips or establishes stiction depending on subtle state variations.

STRIDE represents the unstructured component as a conditional stochastic process that captures the distribution of residual forces given the system state using CFM. It learns a continuos transport vector field that maps simple noise to samples from the conditional distribution,

$$\mathbf{F}_{\text{ext}} \sim p_\phi(\mathbf{F}_{\text{ext}} \mid \mathbf{c}, \mathbf{z}), \tag{7}$$

enabling the model to represent multi-modal interaction behavior.

Specifically, we sample latent noise $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and learn a context-conditioned vector field $v_\phi(\mathbf{z}, t \mid \mathbf{c})$, where the context is $\mathbf{c} = (\mathbf{q}, \dot{\mathbf{q}}, \boldsymbol{\tau})$. The residual force sample is obtained by integrating the learned flow:

$$\boldsymbol{\epsilon}_{\text{CFM}} = \mathbf{z}_0 + \int_0^1 v_\phi(\mathbf{z}_t, t \mid \mathbf{c})\, dt. \tag{8}$$

The generative residual formulation is particularly beneficial in regimes where interaction dynamics exhibit branching behavior, such as near contact transitions or frictional regime changes. Sampling-based prediction with CFM provides a more faithful representation of the underlying uncertainty while remaining compatible with model-based control.

A. Joint Optimization

A key feature of STRIDE is that the LNN and CFM components are trained jointly under a single supervised objective on observed accelerations. Given a dataset $\mathcal{D} = \{(\mathbf{c}, \ddot{\mathbf{q}})\}$, we minimize

$$\mathcal{J}(\theta, \phi) = \mathbb{E}_{(\mathbf{c}, \ddot{\mathbf{q}}) \sim \mathcal{D}, \mathbf{z} \sim \mathcal{N}}\Big[\Big\|\ddot{\mathbf{q}} - \Big(f_{\text{LNN}}(\mathbf{c}; \theta) + \mathbf{M}_\theta^{-1}(\mathbf{q})\boldsymbol{\epsilon}_{\text{CFM}}(\mathbf{c}, \mathbf{z}; \phi)\Big)\Big\|^2\Big]. \tag{9}$$

Optimizing both components jointly encourages an implicit division of labor: the LNN captures the low-variance

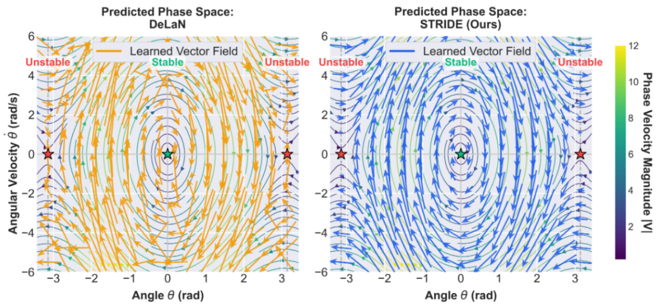

Fig. 1. Phase portrait on the 1-DoF pendulum near the unstable upright equilibrium. STRIDE preserves physically consistent trajectories and captures multi-modal next-state evolution, while deterministic baselines show averaging bias and drift in sensitive regions.

structured dynamics, while the CFM residual models stochastic variability.

IV. Experiments and Results

We evaluate whether the proposed STRIDE formulation delivers measurable gains in prediction accuracy, control performance through (1) long-horizon predictive accuracy and stability under compounding rollout errors, (2) accuracy in modeling non-conservative effects, including contact impulses and impact-induced discontinuities. We begin with targeted experiments designed to validate the key claims introduced in the methodology.

A. Empirical Validation of Design Choices:

Residual Allocation: A potential concern is whether the residual component absorbs dynamics that should be captured by the structured LNN. In a bouncing pendulum experiment comparing f_{LNN} with the residuals, we observe a clear separation of responsibilities: during smooth rigid-body motion (e.g., flight or free swing), the generative residual contributes only 6.5% of the predicted acceleration magnitude. In contrast, during contact transitions and boundary impacts—where structured Lagrangian assumptions degrade—the residual scales to 78.8%, absorbing high-frequency discontinuities without corrupting the global mass matrix.

Inference and Accuracy Comparison to LNN + Diffusion model: We compare STRIDE's CFM-based residual generation with the diffusion residual used in an LNN + Diffusion baseline. CFM achieves significantly higher inference frequency across 2–100 Number of Function Evaluations (NFEs), with especially strong gains at low NFEs. At 5 NFEs, STRIDE attains more than double the inference frequency of the diffusion-based model. Additionally, the flow-based approach reaches the DeLaN rollout error level using only 25% of the NFEs required by diffusion. Overall, STRIDE demonstrates a favourable accuracy–efficiency trade-off, making it a feasible model for real-world deployment.

Behavior in Sensitive Dynamical Regimes: We analyze a 1-DoF pendulum to assess phase portrait consistency in sensitive regions (Fig. 1). Deterministic models (DeLAN) deviates near the unstable equilibrium. In contrast, STRIDE yields a coherent phase portrait aligned with ground truth, preserving elliptical orbits around the stable equilibrium and the saddle structure near the upright configuration. This suggests that, beyond capturing non conservative dynamics, modeling stochastic residuals mitigates averaging effects while preserving the system's qualitative topology.

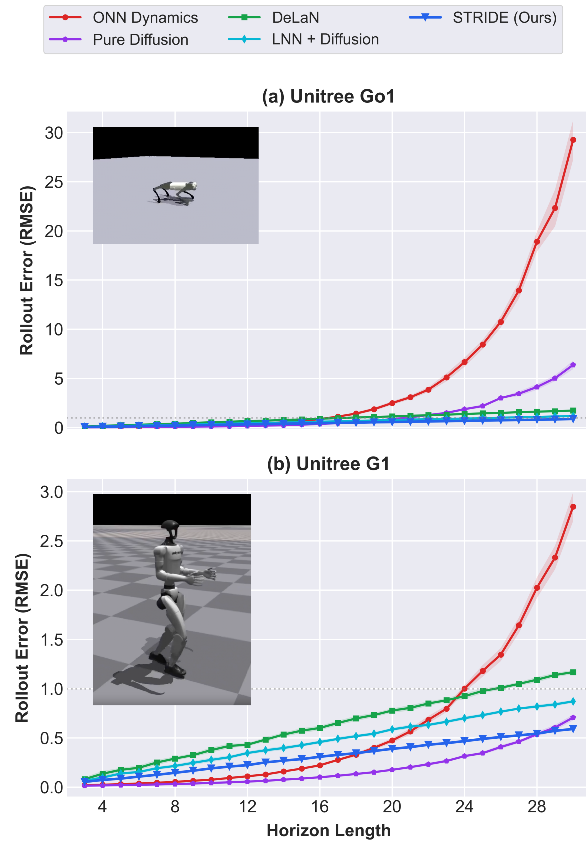

Fig. 2. Long-horizon rollout error on complex legged systems (Unitree Go1 and Unitree G1). The ONN baseline exhibits rapid, near-exponential error growth, while the DeLaN reduces error growth to approximately linear. STRIDE further reduces long-horizon drift by capturing stochastic, contact-induced variability.

B. MPC Integration and Deployment

To evaluate the practical utility of the learned dynamics model, we embed STRIDE within a sampling-based MPC framework, specifically Model Predictive Path Integral (MPPI) control. We integrate STRIDE into the Dreamer–MPC control architecture of [17], where reinforcement learning is used to learn a warm-start policy along with reward, and value models, while MPPI performs online trajectory optimization at deployment. This setup provides a stringent evaluation: the receding-horizon MPC loop repeatedly exposes long-horizon prediction errors, the modular design enables controlled swapping of dynamics models for fair comparison, and the contact-rich nature of legged locomotion stresses modeling assumptions.

C. Long-Horizon Prediction and Stability

We evaluate long-horizon predictive accuracy via multi-step rollouts for both Unitree Go1 and Unitree G1 robots in

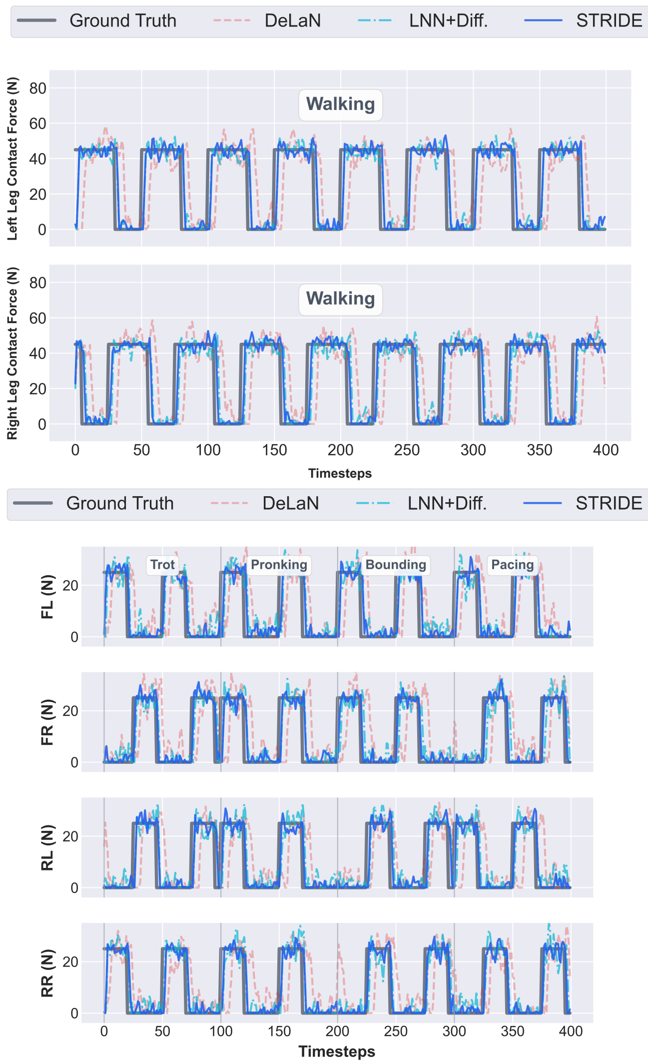

Fig. 3. Contact force prediction on the Unitree G1 humanoid and Unitree Go1 quadruped during walking. Predicted vertical ground reaction forces are compared against ground truth for each leg. STRIDE closely tracks the timing and magnitude of stance–swing transitions, preserving sharp impact discontinuities and reducing force smoothing observed in deterministic baselines (DeLaN and LNN+Diff). For Unitree Go1, comparison is shown across multiple gaits (trot, pronk, bound, and pace).

simulation and report the cumulative root-mean-square error (RMSE) over a horizon of $H = 30$ (Fig. 2). Table I reports the cumulative normalized RMSE over the 30-step horizon. STRIDE reduces rollout error by 83% relative to the ONN baseline on Unitree Go1 (53% on Unitree G1), and achieves a further 19% (Unitree Go1) and 21% (Unitree G1) reduction compared to the strongest structured generative baseline (LNN + Diffusion), representing a substantial improvement over models that already incorporate physical priors. Interestingly, the Pure Diffusion baseline performs slightly better on the Unitree G1 humanoid, likely because direct observation-space modeling can more readily absorb high-dimensional sensing noise, whereas structure-constrained models prioritize physically consistent acceleration prediction.

TABLE I
PERFORMANCE COMPARISON. (EVALUATED AT HORIZON = 30)

Robot	Model	State RMSE ↓	Force Error (N) ↓
Go1	Plain MLP	5.489	NA
	Pure Diffusion	3.521	NA
	DeLaN	1.326	9.4
	LNN + Diffusion	1.154	8.8
	STRIDE (Ours)	**0.932**	**6.7**
G1	Plain MLP	1.204	NA
	Pure Diffusion	0.314	NA
	DeLaN	1.023	17.1
	LNN + Diffusion	0.918	15.4
	STRIDE (Ours)	**0.289**	**12.1**

D. Contact and Force Prediction

We evaluate contact modeling accuracy by comparing predicted ground reaction forces with ground-truth measurements (Fig. 3). Since the plotted forces correspond to the non-conservative term F_{ext}, they are directly available only for LNN-based models; thus, force-level comparisons are limited to structured baselines.

Beyond reducing aggregate RMSE, STRIDE captures sharp impact discontinuities and swing–stance transitions, preserving both timing and magnitude of impulsive forces (Fig. 3). The resulting force prediction error is 12% of nominal contact force on Unitree Go1 and 7% on Unitree G1, indicating accurate modeling relative to physical force scales. Improvements are consistent across legs and across four quadruped gaits, suggesting generalization across contact configurations rather than overfitting to specific locomotion modes.

Overall, STRIDE achieves substantially lower force prediction error than all structured baselines (Table I), reducing error by 30% relative to DeLaN on both quadruped and humanoid systems. This improves contact-aware planning and closed-loop stability on hardware. Consistent gains on the higher-dimensional humanoid further indicate robust scalability to more complex platforms.

V. CONCLUSION

We presented STRIDE, a dynamics modeling framework that separates conservative rigid-body mechanics from stochastic non-conservative interaction effects. The structured component leverages a Lagrangian Neural Network to preserve physical inductive bias, while residual forces are modeled via Conditional Flow Matching to efficiently capture multi-modal variability. Across systems of increasing complexity, including the Unitree Go1 quadruped and Unitree G1 humanoid, STRIDE improves long-horizon prediction accuracy and contact force prediction, enabling more reliable model-based control, particularly under frequent contact transitions.

Future work will focus on integrating STRIDE with uncertainty-aware planning and risk-sensitive MPC, improving robustness under distribution shift, and enabling online adaptation. Extending the residual model to incorporate visual inputs for perception-driven, environment-aware interaction modeling is a promising direction.

References

[1] Y. Hu, J. Liu, A. Spielberg, J. B. Tenenbaum, W. T. Freeman, J. Wu, D. Rus, and W. Matusik, "Chainqueen: A real-time differentiable physical simulator for soft robotics," in *2019 International conference on robotics and automation (ICRA)*, pp. 6265–6271, IEEE, 2019.

[2] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*. Springer, 2010.

[3] R. Featherstone, "Rigid body dynamics algorithms," *Springer*, 2014.

[4] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," *IROS*, 2012.

[5] M. P. Deisenroth and C. E. Rasmussen, "Pilco: A model-based and data-efficient approach to policy search," *ICML*, 2011.

[6] G. Williams, A. Aldrich, and E. A. Theodorou, "Information-theoretic model predictive control: Theory and applications to autonomous driving," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1603–1622, 2017.

[7] N. Gruver, M. A. Finzi, S. D. Stanton, and A. G. Wilson, "Deconstructing the inductive biases of hamiltonian neural networks," in *International Conference on Learning Representations*, 2022.

[8] M. Lutter, C. Ritter, and J. Peters, "Deep lagrangian networks: Using physics as an inductive bias for deep learning," in *International Conference on Learning Representations (ICLR)*, 2019.

[9] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho, "Lagrangian neural networks," in *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

[10] Y. D. Zhong, B. Dey, and A. Chakraborty, "Extending lagrangian and hamiltonian neural networks with differentiable contact models," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21910–21922, 2021.

[11] L. Schulze, J. Peters, and O. Arenz, "Context-aware deep lagrangian networks for model predictive control," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6939–6946, IEEE, 2025.

[12] P. Kotecha, A. Shirwatkar, and S. Kolathaya, "Investigating lagrangian neural networks for infinite horizon planning in quadrupedal locomotion," *arXiv preprint arXiv:2506.16079*, 2025.

[13] C. Pan, Z. Yi, G. Shi, and G. Qu, "Model-based diffusion for trajectory optimization," *Advances in Neural Information Processing Systems*, vol. 37, pp. 57914–57943, 2024.

[14] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 6840–6851, 2020.

[15] X. Xu, H. Ha, and S. Song, "Dynamics-guided diffusion model for sensor-less robot manipulator design," in *Conference on Robot Learning*, pp. 4446–4462, PMLR, 2025.

[16] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," in *The Eleventh International Conference on Learning Representations*, 2023.

[17] A. Shirwatkar, N. Saxena, K. Chandra, and S. Kolathaya, "Piploco: A proprioceptive infinite horizon planning framework for quadrupedal robot locomotion," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11198–11204, IEEE, 2025.