

TACTIC: Tactile and Vision Conditioned Contact Centric Control for Whole-Arm Manipulation

Rishabh Madan¹, Angchen Xie², Samantha Saak¹, Andres Blanco¹, Dohyeok Lee¹,
Sarah Grace Brown¹, Yunting Yan¹, Mark Zolotas³, Jose Barreiros³, Tapomayukh Bhattacharjee¹
¹Cornell University ²Carnegie Mellon University ³Toyota Research Institute



Fig. 1: **TACTIC** enables robots to perform contact-rich manipulation by utilizing the entire arm surface. While end-effector-centric control is often infeasible and unsafe under high payloads, TACTIC fuses multimodal observations such as RGB-D, proximity masks, and distributed tactile sensing into a contact-centric representation to enable whole-arm manipulation. We demonstrate its effectiveness on manipulation tasks, including side rollover and limb repositioning on a life-size manikin.

Abstract—Whole-arm manipulation involves direct contact with the environment while the robot completes a task by distributing contact across multiple links as contacts form, slide, and break. This setting breaks common implicit assumptions in many learning-based manipulation pipelines: arm configuration tightly couples motion and contact forces, contact state is partially observed under occlusion, and purely learned rollouts can become physically inconsistent under distribution shift because many multi-link contact configurations are sparsely represented in the data. To address this, we propose TACTIC (Tactile and Vision Conditioned Contact-Centric Control), a receding-horizon controller for whole-arm manipulation. TACTIC uses a contact-centric hybrid predictive model that combines RGB-D, distributed tactile sensing, and a compact 2D proximity representation. The model couples a learned, action-conditioned latent dynamics model with analytical kinematics through contact

Jacobians, enabling rollouts of future contact configurations and interaction forces. TACTIC integrates these rollouts into a sampling-based MPC planner with contact-aware action sampling: contact Jacobian-based projections steer sampled action sequences toward force-modulating directions, and objectives defined over predicted proximity and interaction forces trade task progress against whole-arm force regulation. We evaluate TACTIC in simulation against state-of-the-art model-based and model-free methods, and perform ablations that isolate the contribution of each design choice. Across experiments, TACTIC consistently outperforms other methods. We further demonstrate real-world performance on a robot with distributed tactile sensing across three whole-arm manipulation tasks that require multi-contact trajectories: turning over and repositioning a manikin, and goal-reaching in a 3D dynamic maze. Website: emprise.cs.cornell.edu/tactic.

I. INTRODUCTION

Recent progress in robot learning spanning vision-language-action (VLA) models [1], diffusion-based policies [2], and world models [3] has improved performance across a wide range of manipulation tasks [4]. By leveraging large datasets, these methods acquire priors over scenes and skills and often yield desired behavior with minimal task-specific engineering. However, a gap remains: these approaches struggle in *contact-rich manipulation* where success requires reasoning about and regulating interaction forces. This gap is especially pronounced in *whole-arm manipulation*, where the robot deliberately uses multiple arm links to make, break, and regulate contact while executing a task.

Whole-arm contact is fundamental to human manipulation: people use the forearm and upper arm to brace, hold, guide, and reposition objects and bodies, distributing forces over a larger area to improve stability and comfort [5]. Robots require similar capabilities in contact-rich settings such as assistive caregiving and physical human-robot interaction (pHRI). These challenges also arise when manipulating large, unwieldy, or deformable objects in cluttered environments, where contacts form and redistribute along the arm. Whole-arm manipulation introduces two key challenges. First, trajectory generation is hard because arm configuration tightly couples motion and forces: it determines which links make contact with the environment, where contacts form and break, and how forces evolve over time. Second, data is scarce for the regimes that matter most for whole-arm manipulation: diverse multi-contact configurations and force-regulating behaviors are difficult to collect and remain sparsely represented in existing datasets.

Existing approaches often address only a subset of these challenges, typically emphasizing either learned models or analytical structure. Purely data-driven models can be brittle because widely used datasets are dominated by end-effector-centric interactions and lack the diversity of multi-link contact configurations needed to generalize to whole-arm behaviors. Even with tactile supervision, learned predictors can produce physically inconsistent contact and force estimates (e.g., unrealistic force magnitudes) under distribution shift. Conversely, approaches that rely solely on analytical models [6, 7] provide explicit mechanisms for safety but often fail to capture the multimodal, partially observed interactions that arise in contact-rich whole-arm manipulation. Relying on either paradigm alone can therefore yield systems that either capture rich interactions but struggle to regulate safety-related objectives under distribution shift, or enforce safety without generating complex whole-arm behaviors.

Reliable whole-arm manipulation requires reasoning about simultaneous contacts distributed along the arm. Because contact locations and interaction forces are difficult to infer from joint torques alone, we use distributed tactile sensing to observe contact directly. Yet sensing contact is not enough; the controller must represent, predict, and act on contact information to balance task progress with whole-arm force

regulation. We use receding-horizon sampling-based MPC as our control framework, for two reasons: receding-horizon control enables continual replanning as contacts form, break, and shift across the arm, while sampling-based optimization can evaluate diverse candidate actions under nonconvex and contact-dependent dynamics. Within this framework, we ground our approach in three design principles. First, the state should be contact-centric and multimodal: we fuse RGB-D with distributed tactile sensing and a compact 2D proximity representation that summarizes where the arm is near or in contact with the scene (including a person) and how this region evolves over time. Second, planning should use contact-aware action sampling: in constrained multi-contact settings, standard sampling schemes (e.g., MPPI, CEM) are inefficient at discovering force-modulating actions. We use tactile sensor-derived contact locations to compute contact Jacobians, which allow us to bias sampled action sequences toward directions that can regulate interaction forces. Third, prediction and evaluation should be hybrid and contact-aware: we couple a learned, action-conditioned latent dynamics model with analytical kinematics via contact Jacobians to roll out contact-relevant dynamics and interaction forces, and we evaluate candidate trajectories with objectives defined over predicted proximity and forces to trade task progress against whole-arm force regulation.

Built on these principles, we propose **TACTIC** (*Tactile and Vision Conditioned Contact-Centric Control*), a receding-horizon controller for whole-arm manipulation. TACTIC uses a *contact-centric* state representation by coupling RGB-D, distributed tactile sensing, and a proximity map around the arm. To explore efficiently under multi-contact constraints, TACTIC uses sampling-based MPC with *contact-aware action sampling*, biasing candidate action sequences toward force-modulating directions via contact Jacobian projections. For prediction and trajectory scoring, TACTIC couples an action-conditioned latent dynamics model (trained on the contact-centric state) with analytical kinematics through contact Jacobians to predict the evolution of proximity and interaction forces, and optimizes objectives defined over these predictions. We evaluate TACTIC through simulation studies that compare against competitive baselines and include ablations to quantify the impact of each design choice. We further demonstrate real-world performance on a robot with distributed tactile sensing in two contact-rich whole-arm manipulation tasks involving a life-size manikin, side rollover and limb repositioning, and in a 3D dynamic maze goal-reaching task that requires complex multi-contact trajectories. Finally, we develop and release open hardware for a mirrored Kinova Gen3-7DoF [8] teleoperation interface and a WiFi-based distributed piezoresistive tactile sensing suite for the Gen3 arm.

The main contributions of this paper are as follows:

- C1.** *Contact-centric* state representation for whole-arm manipulation that fuses RGB-D, distributed tactile sensing, and a compact 2D proximity map.
- C2.** *Contact-aware action sampling* for sampling-based MPC that uses contact Jacobian projections to bias exploration

toward force-modulating directions in constrained multi-contact settings.

- C3.** *Hybrid predictive model* that couples learned, action-conditioned latent dynamics with analytical kinematics via contact Jacobians to predict contact dynamics and forces, and scores trajectories with objectives defined over predicted proximity and interaction forces.
- C4.** Simulation and real-robot experiments with ablations, demonstrating improved performance relative to purely learned and purely analytical baselines across multi-contact tasks.
- C5.** Open hardware for a Kinova Gen3 exoskeleton and a WiFi-based distributed piezoresistive tactile sensing suite, including design files, assembly instructions, and code.

II. RELATED WORK

Whole-arm manipulation. Recent research in whole-arm manipulation (WAM) leverages distributed tactile skins to localize and regulate contact [9–12]. These capabilities have been applied to manipulating heavy or bulky objects [13, 14], navigating clutter [6], and providing physical [15] and social [16] assistance. Prior studies show that allowing whole-arm contact improves reachability in cluttered environments [6]. Furthermore, example-guided reinforcement learning can acquire robust whole-body manipulation strategies for large objects on compliant humanoid hardware with tactile sensing [14]. Despite these advances, many WAM systems are not designed for whole-arm pHRI, where the robot must infer task-relevant contact state from multimodal observations (vision, proprioception, and tactile) under occlusion and regulate interaction forces online. In contrast, our approach leverages multimodal sensing to adapt contact-rich whole-arm behaviors and incorporates explicit force-related objectives within MPC to support safer physical interaction.

Visuotactile representations for manipulation. Contact-rich manipulation often demands feedback beyond vision due to occlusion, compliance, and rapidly changing contact modes. Visuotactile representations, therefore, fuse vision and touch to infer task-relevant contact state and improve robustness during interaction (e.g., self-supervised multimodal representations for contact-rich tasks and visuotactile policies that adapt after contact) [17–19]. These representations have also been paired with model-based control to close the loop on contact, including tactile servoing and predictive control frameworks that use learned dynamics from touch (and, in some cases, vision) to plan under contact uncertainty [20, 21]. In parallel, model-based learning for contact-rich manipulation learns task-specific contact models or action-conditioned world models and embeds them into planning to improve predictive accuracy under contact [3, 22–24]. However, many such pipelines remain primarily vision-centric and do not explicitly exploit distributed tactile supervision or analytical kinematic structure, which can be valuable for grounding contact reasoning and enforcing force-aware objectives in pHRI. In contrast, our approach uses a contact-centric visuotactile representation and integrates it with analytical kinematics via

Symbol	Meaning	Symbol	Meaning
n	arm DoF	H	MPC horizon
$\mathbf{q}_t, \dot{\mathbf{q}}_t$	joint pos./vel.	N	MPPI rollouts
$\mathbf{u}_t \in \mathbb{R}^n$	cmd joint vel. (input)	$\boldsymbol{\mu}_{t+i}$	MPPI mean at $t+i$
$\mathcal{U}$	admissible actions (e.g., limits)	Σ	MPPI covariance
		$\mathcal{J}(\mathbf{u}; o_{1:t}, o_g)$	MPC objective
o_t	multimodal observation	$\mathbf{f}_t \in \mathbb{R}^{N_s}$	tactel forces (N_s sensors)
o_g	goal observation	$\psi_{\text{contact}}(\hat{\mathbf{z}}_{t+i})$	decoded binary contacts
o_t^{cc}	contact-centric obs. from o_t	F_{safe}	safe-force limit
M_t	proximity mask	$\mathcal{K}_t$	active contacts
$\mathbf{z}_t, \hat{\mathbf{z}}_{t+i}$	latent / predicted latent	$(\mathcal{K}_t = K_t)$	
$\phi(\cdot)$	encoder	$\mathbf{J}_t^n$	contact Jacobian (normal)
$\mathbf{F}_l$	latent dynamics model	$f_{k,t}$	force magnitude (contact k , time t)
$\mathbf{F}_k$	analytical kinematics model	F_{sat}	saturation force

TABLE I: Glossary of notation used in this paper.

joint state propagation and contact Jacobians for contact-aware prediction and planning.

III. PROBLEM FORMULATION

We consider whole-arm manipulation where the robot accomplishes a task by moving its arm while making, breaking, and regulating contact along its links. The robot is an n -DoF manipulator with joint configuration $\mathbf{q}_t \in \mathbb{R}^n$, joint velocity $\dot{\mathbf{q}}_t \in \mathbb{R}^n$, and control input given by commanded joint velocity $\mathbf{u}_t \in \mathbb{R}^n$. The robot receives a multimodal observation $o_t = (I_t, D_t, \mathbf{f}_t, \mathbf{q}_t, \dot{\mathbf{q}}_t)$, where I_t represents an RGB image, D_t represents an aligned depth map, and $\mathbf{f}_t \in \mathbb{R}^{N_s}$ are distributed tactile readings over N_s taxels (tactile pixels) along the arm. We are given a goal observation o_g and safety constraints, including maximum allowable contact force bounded by a threshold F_{safe} . Our objective is to reach the goal while respecting these safety limits.

We use *sampling-based* receding-horizon MPC, and instantiate the online optimizer with Model-Predictive Path Integral (MPPI) [25]. At each time step t , MPPI approximates the solution of a finite-horizon constrained problem by sampling candidate action sequences, evaluating their costs, and updating the nominal sequence toward lower-cost rollouts. We then execute only the first action and replan at the next step with updated observations:

$$\mathbf{u}_{t:t+H-1}^* = \underset{\mathbf{u}_{t:t+H-1} \in \mathcal{U}^H}{\operatorname{argmin}} \mathcal{J}(\mathbf{u}_{t:t+H-1}; o_{1:t}, o_g) \quad (1)$$

IV. TACTIC FOR WHOLE-ARM MANIPULATION

This section outlines the core design decisions of TACTIC as a sampling-based controller for whole-arm manipulation.

A. Building a Contact-Centric Representation

TACTIC learns an action-conditioned dynamics model from vision, distributed force data, and proprioception. From the raw observation o_t in Eq. 1, we compute a proximity mask M_t and construct the augmented observation $o_t^{\text{cc}} = (I_t, D_t, M_t, \mathbf{f}_t, \mathbf{q}_t, \dot{\mathbf{q}}_t)$.

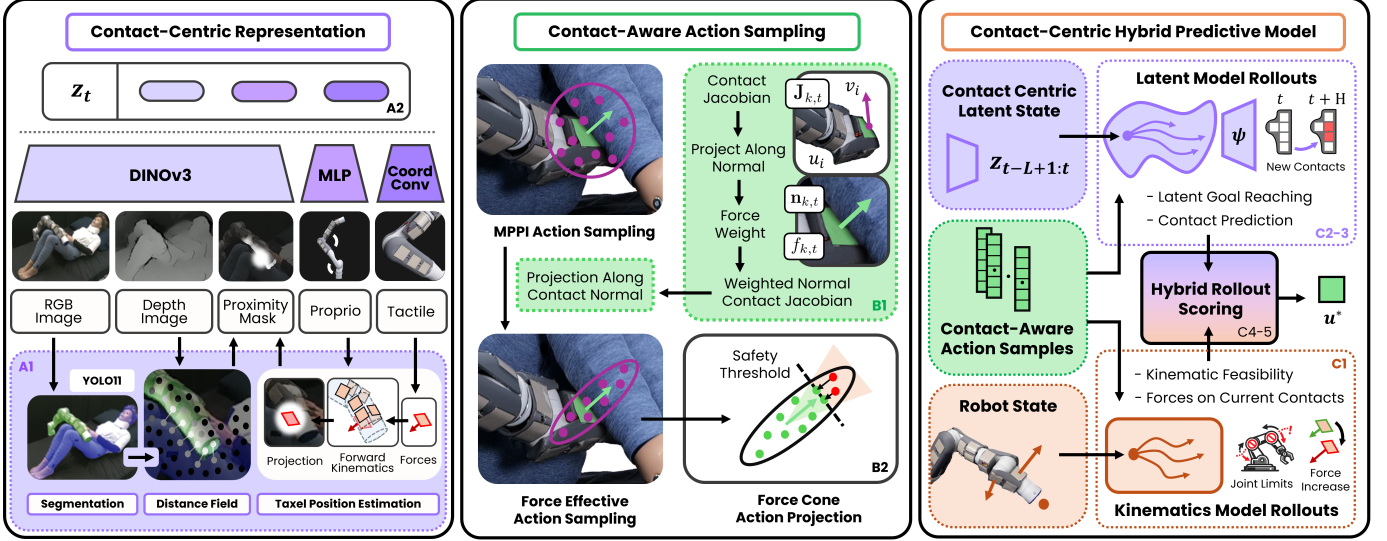


Fig. 2: Overview of TACTIC. **A. Contact-Centric Representation** encodes multimodal history into a latent state z_t . **B. Contact-Aware Action Sampling** utilizes tactile-derived contact Jacobians ($J_{k,t}$) to bias MPPI exploration toward force-modulating directions. **C. Contact-Centric Hybrid Predictive Model** evaluates candidate sequences by combining an analytical kinematics branch for physical consistency with a learned latent dynamics model for predicting contact and force evolution.

1. Proximity mask M_t . Tactile readings are spatially sparse, available only upon contact, and localize contact on the arm rather than in the scene. Yet many tasks require dense, anticipatory reasoning about *where in the scene* contact is occurring or likely to occur next. We define the proximity mask M_t as an image-aligned 2D representation of robot-object proximity that grounds impending and active contact in the scene frame, highlighting arm surface regions near or contacting the object. First, we segment the RGB image into robot (Ω_r) and object (Ω_o) masks using finetuned YOLOv11 [26]. Using camera intrinsics, we unproject pixels in Ω_r and Ω_o into 3D point sets P_r and P_h , respectively, then compute each robot point’s distance to the closest object point using kd-tree nearest-neighbor queries [27]. We map these distances to pixel intensity (near $\rightarrow$ bright, far $\rightarrow$ dark) to obtain a dense visual proximity field. Finally, we fuse tactile data by rendering a binary taxel indicator image M_t^{tax} , initialized to zero. For each taxel with force exceeding $f_{\min}$, we project its 3D pose via forward kinematics onto the image and render a unit-valued blob of radius r at the projected location. The final M_t is the per-pixel maximum of the visual proximity field and M_t^{tax} , providing an explicit, visually grounded signal for both impending and active contact. See Appendix B1 for implementation details.

2. Encoding multimodal observations. We encode o_t^{cc} into a latent z_t using modality-specific encoders. We use a DINOv3 backbone [28] and apply it to I_t , D_t , and M_t using a shared encoder in a batched forward pass, with D_t and M_t replicated to three channels. DINOv3’s robust spatial features transfer well to depth and mask inputs. For force sensing, we treat f_t as a vector of normal forces under a fixed taxel ordering along the robot arm. Contact events in our dataset are spatially sparse. To efficiently detect these local features without significantly overparameterizing the model, we use

a lightweight convolutional network with 1D CoordConv [29] layers. By concatenating the taxel index as an explicit position channel, this architecture enables the network to learn location-dependent contact features. For proprioception, we use a lightweight MLP that jointly embeds $(q_t, \dot{q}_t)$. To avoid angle wrap-around, we represent each joint angle using a sin-cos mapping before passing to the proprioceptive encoder. The latent dynamics model (Sec. IV-C) conditions on a history of L latent states, $z_{t-L+1:t}$.

B. Contact-Aware Action Sampling

We use MPPI, a sampling-based MPC algorithm, to generate command actions for the robot. At each MPC time step t , MPPI samples N action sequences over a horizon of length H . Standard MPPI samples perturbations $\delta_{t+i}^{(j)} \sim \mathcal{N}(0, \Sigma)$ for rollout $j \in \{1, \dots, N\}$ and horizon index $i \in \{0, \dots, H-1\}$, and sample action sequence $u_{t:t+H-1}^{(j)}$. We build on STORM [30] as our base MPPI sampler, which uses low-discrepancy Halton sequences and spline-based sample smoothing, and adapts the mean and covariance parameters via cost-weighted sample moments.

In whole-arm manipulation, the set of active contacts can change from step to step. If contact handling enters only through the scalar rollout cost, the sampler can adapt slowly: covariance updates lag behind the current contact interaction, especially with a smaller particle budget. We therefore incorporate contact information directly into the sampler by shaping perturbations per-MPC-step, providing minimal-lag, contact-conditioned exploration toward the *contact normal* subspace.

Although STORM is formulated over joint accelerations, we use it as an *auxiliary sampler* to generate smooth perturbation sequences while keeping our MPPI policy in joint-velocity space. This allows for smoother perturbations in joint-velocity space. Concretely, for each rollout j , we draw

acceleration noise $\boldsymbol{\eta}_{t:t+H-1}^{(j)}$ using Halton sampling with B-spline smoothing and integrate it over the horizon to obtain joint-velocity perturbations $\boldsymbol{\delta}_{t:t+H-1}^{(j)}$ (with time step Δt), $\boldsymbol{\delta}_{t+i}^{(j)} = \Delta t \sum_{\ell=0}^i \boldsymbol{\eta}_{t+\ell}^{(j)}$, and apply $\mathbf{u}_{t+i}^{(j)} = \boldsymbol{\mu}_{t+i} + \boldsymbol{\delta}_{t+i}^{(j)}$ to generate joint velocity action $\mathbf{u}_{t+i}^{(j)}$ around nominal mean $\boldsymbol{\mu}_{t+i}$.

1. Weighted normal contact Jacobian. Let $\mathcal{K}_t$ be the active contact set at time t . For each $k \in \mathcal{K}_t$, we compute the contact Jacobian $\mathbf{J}_{k,t} \in \mathbb{R}^{3 \times n}$ and a unit contact normal $\mathbf{n}_{k,t} \in \mathbb{R}^3$. The corresponding normal Jacobian row $\mathbf{J}_{k,t}^n = \mathbf{n}_{k,t}^\top \mathbf{J}_{k,t}$ maps joint velocities to normal velocity at contact. Stacking rows gives $\mathbf{J}_t^n \in \mathbb{R}^{K_t \times n}$, and the predicted normal velocity under rollout j is $v_{n,k,t+i}^{(j)} = \mathbf{J}_{k,t}^n \mathbf{u}_{t+i}^{(j)}$. Intuitively, directions in joint-velocity space that lie in $\text{row}(\mathbf{J}_t^n)$ change normal motion at the current contacts; directions in its nullspace do not.

Not all “active contacts” are equally important. A light brush or noisy taxel should not dominate sampling. Thus, we weight each contact by its measured force (with saturation) before constructing the contact-normal Jacobian:

$$\rho_{k,t} = \text{sat}\left(\frac{\|\mathbf{f}_{k,t}\|}{F_{\text{sat}}}\right), \quad \widetilde{\mathbf{J}}_t^n = \text{diag}(\sqrt{\rho_{1,t}}, \dots, \sqrt{\rho_{K_t,t}}) \mathbf{J}_t^n, \quad (2)$$

where $\text{sat}(x) = \min(1, \max(0, x))$. The resulting $\widetilde{\mathbf{J}}_t^n$ maps joint velocities to normal contact velocities, and its row space spans the joint directions that can change normal motion at the current contacts.

2. Contact-conditioned shaping. We define a contact-normal filter:

$$\mathbf{G}_t^{\text{force}} = \widetilde{\mathbf{J}}_t^{n\top} \left(\widetilde{\mathbf{J}}_t^n \widetilde{\mathbf{J}}_t^{n\top} + \varepsilon \mathbf{I}_{K_t} \right)^{-1} \widetilde{\mathbf{J}}_t^n, \quad (3)$$

$$\mathbf{G}_t^{\text{null}} = \mathbf{I}_n - \mathbf{G}_t^{\text{force}}.$$

$\mathbf{G}_t^{\text{force}}$ extracts the component of a perturbation aligned with $\text{row}(\widetilde{\mathbf{J}}_t^n)$, the directions that change normal velocity at active contacts, while $\mathbf{G}_t^{\text{null}}$ retains the complementary component. The ridge term $\varepsilon \mathbf{I}_{K_t}$ stabilizes the inverse when contact normals are nearly collinear. As $\varepsilon \rightarrow 0$, $\mathbf{G}_t^{\text{force}}$ recovers the orthogonal projector onto $\text{row}(\widetilde{\mathbf{J}}_t^n)$ and $\mathbf{G}_t^{\text{null}}$ recovers the projector onto its nullspace.

Assuming the contact configuration remains valid over the short planning horizon, we shape the perturbation sequences using the above operators. Let $\boldsymbol{\delta}_{t+i}^{(j)}$ denote the base perturbation. We construct

$$\mathbf{S}_t = \sigma_{\text{force,eff}} \mathbf{G}_t^{\text{force}} + \sigma_{\text{null,eff}} \mathbf{G}_t^{\text{null}}, \quad (4)$$

$$\sigma_{\bullet,\text{eff}} = \sqrt{\sigma_{\bullet}^2 + \sigma_{\min}^2}, \quad \bullet \in \{\text{force}, \text{null}\},$$

$\mathbf{S}_t$ is a contact-conditioned scaling that amplifies perturbations along contact-normal directions relative to nullspace directions. We apply it to obtain contact-conditioned perturbations, $\boldsymbol{\delta}_{t+i}^{(j),\text{shaped}} = \mathbf{S}_t \boldsymbol{\delta}_{t+i}^{(j)}$. Setting $\sigma_{\text{force}} > \sigma_{\text{null}}$ biases exploration toward directions that affect contact interaction; $\sigma_{\min}$ ensures a nonzero scale in both subspaces.

Finally, to avoid sampling actions that drive motion into contact where forces are high, we project each sampled action sequence onto a safety cone defined by the active normals

above a force threshold. Let $\mathcal{K}_t^{\text{viol}} = \{k \in \mathcal{K}_t \mid f_{k,t} \geq F_{\text{safe}}\}$ and let $\mathbf{J}_{t+i}^{\text{viol}}$ stack rows $\mathbf{J}_{k,t}^n$ for $k \in \mathcal{K}_t^{\text{viol}}$. For each rollout j and horizon step i , we solve

$$\boldsymbol{\delta}_{t+i}^{(j),\text{final}} = \underset{\mathbf{y} \in \mathbb{R}^n}{\text{argmin}} \left\| \mathbf{y} - \boldsymbol{\delta}_{t+i}^{(j),\text{shaped}} \right\|_2^2 \quad (5)$$

s.t. $\mathbf{J}_{t+i}^{\text{viol}}(\boldsymbol{\mu}_{t+i} + \mathbf{y}) \geq \mathbf{0}$, where rows are oriented such that $\mathbf{J}_{t+i}^{\text{viol}} \mathbf{u}_t \geq \mathbf{0}$ corresponds to commanded normal motion away from contact. The obtained action is thus $\mathbf{u}_{t+i}^{(j)} = \boldsymbol{\mu}_{t+i} + \boldsymbol{\delta}_{t+i}^{(j),\text{final}}$. See Appendix C for implementation details.

C. Contact-Centric Hybrid Predictive Model

Whole-arm manipulation closely couples joint-space motion and contact: small changes in arm configuration can shift where contact occurs and whether forces increase or decrease. To plan under this coupling, TACTIC uses a *hybrid predictive model* $\mathbf{F} = \{\mathbf{F}_k, \mathbf{F}_l\}$ that separates what we can compute reliably (robot kinematics) from what is difficult to model analytically (contact formation, making and breaking of contacts, and interaction dynamics). The analytical branch provides physically grounded rollouts, while the learned branch predicts the evolution of complex contact-rich interactions from multimodal observations.

1. Analytical kinematics $\mathbf{F}_k$. Given the current joint state and a candidate joint-velocity sequence $\mathbf{u}_{t:t+H-1}$, we roll out deterministic kinematics with first-order integration to obtain $\hat{\mathbf{q}}_{t+i}, \hat{\dot{\mathbf{q}}}_{t+i}$ and compute task-space quantities via forward kinematics, $\hat{\mathbf{e}}_{t+i} = \text{FK}(\hat{\mathbf{q}}_{t+i})$, where $\hat{\mathbf{e}}_{t+i}$ denotes the predicted end-effector pose (or task-space state) at step $t+i$. For contact-aware reasoning, we evaluate the contact Jacobian $\mathbf{J}_{k,t+i}(\hat{\mathbf{q}}_{t+i})$ for each current active contact index $k \in \mathcal{K}_t$, at the predicted configuration $\hat{\mathbf{q}}_{t+i}$ and associate it with the contact normal $\mathbf{n}_{k,t+i}$ yielding the normal Jacobian row $\hat{\mathbf{J}}_{k,t+i}^n = \mathbf{n}_{k,t+i}^\top \mathbf{J}_{k,t+i}$. We collect these predictions into the kinematic rollout $\hat{\mathbf{s}}_{t+i} = (\hat{\mathbf{q}}_{t+i}, \hat{\dot{\mathbf{q}}}_{t+i}, \hat{\mathbf{e}}_{t+i}, \{\mathbf{n}_{k,t}, \hat{\mathbf{J}}_{k,t+i}^n\}_{k \in \mathcal{K}_t})$. The kinematic rollout $\hat{\mathbf{s}}_{t+1:t+H} = \mathbf{F}_k(\mathbf{q}_t, \mathbf{u}_{t:t+H-1})$ provides information needed to (i) reject kinematically infeasible motion and (ii) reason about how candidate actions move the active contacts along the contact normal (via $\hat{\mathbf{J}}_{k,t+i}^n$).

2. Latent dynamics $\mathbf{F}_l$. Kinematics alone cannot predict when contacts will form or break. It also discards the multimodal context needed to perform the task. We therefore learn a latent dynamics model following the JEPA recipe for robotic world models [3, 31]. At each time step, $\mathbf{o}_t^{\text{cc}}$ is encoded into a latent state $\mathbf{z}_t$ using the encoders in Sec. IV-A. We embed $\mathbf{u}_t$ with a lightweight action encoder $\phi_{\text{act}}(\mathbf{u}_t)$. The predictor takes a context window $\mathbf{z}_{t-L+1:t}$ and an action sequence $\mathbf{u}_{t:t+H-1}$ and outputs a latent rollout: $\hat{\mathbf{z}}_{t+1:t+H} = \mathbf{F}_l(\mathbf{z}_{t-L+1:t}, \phi_{\text{act}}(\mathbf{u}_{t:t+H-1}))$. We implement $\mathbf{F}_l$ as a ViT [32] predictor with a frame-causal attention mask [31]. We condition on actions using adaptive layer normalization at each transformer block, following best practices in [33]. We use lightweight decoders ψ_m to map latent rollouts to task-relevant signals. In this work, we decode joint position and binary

contacts for all taxels. These decoders are used for force-based costs in Sec. IV-D.

3. Training objective. We train our models on offline datasets. From each trajectory we sample segments containing a context of length L and a prediction horizon H , encode $\mathbf{z}_t = \phi(o_t^{\text{cc}})$, and predict $\hat{\mathbf{z}}_{t+1:t+H} = \mathbf{F}_l(\mathbf{z}_{t-L+1:t}, \phi_{\text{act}}(\mathbf{u}_{t:t+H-1}))$. Based on findings shared in [33], we minimize a teacher-forced latent consistency loss, augmented with a multistep rollout loss (to reduce drift when the model must consume its own predictions):

$$\mathcal{L} = \underbrace{\sum_{k=1}^H \|\hat{\mathbf{z}}_{t+k} - \text{sg}[\mathbf{z}_{t+k}]\|_2^2}_{\mathcal{L}_{\text{teacher}}} + \underbrace{\sum_{k=2}^{K_{\text{roll}}} \lambda_k \|\hat{\mathbf{z}}_{t+k}^{(k)} - \text{sg}[\mathbf{z}_{t+k}]\|_2^2}_{\mathcal{L}_{\text{multistep}}}$$

where $\hat{\mathbf{z}}_{t+k}^{(k)}$ denotes the k -step autoregressive prediction (up to K_{roll}) obtained by feeding back the model’s own latent predictions, and $\text{sg}[\cdot]$ is stop-gradient on the target latent.

4. Planning with hybrid models. At each MPC step t , we form the contact-centric observation o_t^{cc} (Sec. IV-A) and use MPPI with contact-aware action sampling (Sec. IV-B) to sample N candidate action sequences $\mathbf{u}_{t:t+H-1}^{(j)}$, $j = 1, \dots, N$. For each candidate, we roll out two models in parallel. The analytical branch propagates joint kinematics via $\hat{\mathbf{q}}_{t+1:t+H}^{(j)} = \mathbf{F}_k(\mathbf{q}_t, \mathbf{u}_{t:t+H-1}^{(j)})$ and produces kinematic rollout $\hat{\mathbf{s}}_{t+1:t+H}^{(j)}$ (combining joint states, end-effector states, contact Jacobians for active contacts). Meanwhile, we encode a history of observations into a latent context $\mathbf{z}_{t-L+1:t} = \phi(o_{t-L+1:t}^{\text{cc}})$ and unroll the learned latent dynamics $\hat{\mathbf{z}}_{t+1:t+H}^{(j)} = \mathbf{F}_l(\mathbf{z}_{t-L+1:t}, \mathbf{u}_{t:t+H-1}^{(j)})$ (Sec. IV-C). The resulting rollouts are scored by $\mathcal{J}(\mathbf{u}_{t:t+H-1}^{(j)})$, a weighted combination of different cost functions (discussed in Sec. IV-D). MPPI then updates the nominal sequence toward low-cost rollouts and returns $\mathbf{u}_{t:t+H-1}^*$; we execute the first m actions $\mathbf{u}_{t:t+m-1}^*$ and replan at $t+m$ with updated observations.

5. Two-fidelity rollouts for real-time control. Evaluating $\mathbf{F}_l$ for all N MPPI particles can dominate compute and slow down planning significantly. We use a two-fidelity scheme: a lightweight *draft* predictor $\mathbf{F}_l^{\text{draft}}$ rolls out all particles to obtain approximate costs, then only the top- M candidates ($M \ll N$) are re-scored with the full predictor $\mathbf{F}_l$ before computing MPPI weights. This reduces large ViT evaluations while preserving ranking quality on the best candidates and enabling real-time control. As a safeguard, we monitor the one-step prediction error $\|\hat{\mathbf{z}}_{t+1}^{\text{draft}} - \mathbf{z}_{t+1}\|$ and fall back to the full model when the error exceeds a threshold (see Appendix D).

D. Scoring Hybrid Rollouts

We obtain the best action sequence out of the N sampled sequences by scoring the rollouts generated by our hybrid dynamics model using our cost function $\mathcal{J}$ defined as:

$$\begin{aligned} \mathcal{J}(\mathbf{u}_{t:t+H-1}) = & \sum_{i=1}^H \left(w_{\text{kin}} c_{\text{kin}}(\hat{\mathbf{s}}_{t+i}) \right. \\ & + w_{\text{safety}} c_{\text{force}}(\hat{\mathbf{s}}_{t+i}, \hat{\mathbf{z}}_{t+i}, \mathbf{f}_t) \\ & \left. + w_{\text{goal}} c_{\text{goal}}(\hat{\mathbf{z}}_{t+i}, \mathbf{z}_g) \right), \end{aligned}$$

where c_{kin} penalizes for standard kinematic constraints (self-collision, joint-limit, joint or task-space goals), c_{force} penalizes unsafe predicted forces relative to F_{safe} , combining the learned contact-onset predictor with an analytical force model, and $c_{\text{goal}}(\hat{\mathbf{z}}_{t+i}, \mathbf{z}_g)$ encourages goal-reaching in latent space. We provide the exact cost definitions used in our experiments in Appendix E.

Learned value function for long-horizon tasks To improve latent-space planning on constrained long-horizon tasks, we learn a goal-conditioned value function in latent space from offline data using Implicit Q-Learning (IQL) [34]. For such long-horizon tasks, we assume access to low-dimensional progress variables $\xi_t \in \mathbb{R}^{d_g}$ provided by the dataset (e.g., limb joint angles or body turn angle). At test time, we use a user-specified sequence of *subgoals* $\{\xi_g^{(1)}, \dots, \xi_g^{(G)}\}$, where each $\xi_g^{(g)}$ represents a desired intermediate progress configuration.

We freeze the multimodal encoders (Sec. IV-A) and embed observations and (sub)goals into latents $\mathbf{z}_t$ and $\mathbf{z}_g$. At test time, we augment MPPI with a terminal value term evaluated at the predicted end-of-horizon latent state:

$$\mathcal{J}_{\text{aug}}(\mathbf{u}_{t:t+H-1}) = \mathcal{J}(\mathbf{u}_{t:t+H-1}) - \lambda_V V_{\theta}(\hat{\mathbf{z}}_{t+H}, \mathbf{z}_g),$$

where $\hat{\mathbf{z}}_{t+H}$ is obtained from $\mathbf{F}_l$. Intuitively, V_{θ} estimates reward-to-go beyond the finite horizon, biasing sampling toward action sequences that remain on-task over extended horizons while preserving kinematic feasibility and safety induced by c_{kin} and c_{force} . Training and implementation details are provided in Appendix F.

V. SIMULATION EXPERIMENTS

We first evaluate TACTIC in simulation to isolate the benefits of our contributions by posing the questions below: **Q1:** does having contact-centric representations (**C1**) comprising distributed tactile sensing and the proximity mask improve task success and safety? **Q2:** does contact-aware action sampling (**C2**) improve interaction response, safety, and smoothness? **Q3:** does hybrid modeling (**C3**) by coupling analytical kinematics with learned latent dynamics improve success and efficiency versus either alone?

Environments. We evaluate on 4 environments (Fig. 3):

- **3D Maze Goal Reaching (Maze) RCareWorld [35].** A 3D maze composed of cylinders placed at varying x - y coordinates. Collision-free motion is often infeasible; successful solutions require deliberate whole-arm contact to slide past obstacles and reach a target end-effector pose. We define force thresholds and assess safety via threshold violations.
- **Bed Bathing (Wiping) RCareWorld.** A wheelchair-mounted arm performs a sponge-wiping task on a user lying in a hospital bed. The end-effector trajectory is constrained to



Fig. 3: **Simulation task suite.** We evaluate TACTIC across four simulation environments: the **3D Maze Environment** tests whole-arm navigation through a maze, analogous to navigation in a cluttered environment; **Bed Bathing** tests task execution of wiping a person in bed with incidental whole-arm contact; **Tabletop Reach** tests obstacle-aware reaching across a table; and **Granular Manipulation** tests contact-driven rearrangement of small particles. Whole-arm contact is highlighted in yellow.

follow a wiping motion along the user’s arm. Incidental whole-arm contact occurs as the robot reaches over the bed.

- **Tabletop Reach (Reach)** *Robohive* [36, 37]. A Franka arm reaches across a table to move a cube to a target while avoiding obstacles. This task is used to evaluate hybrid modeling with pretrained world models.
- **Granular Manipulation (Granular)** *FleX* [38]. An xArm manipulates particles on a tabletop and must transform an initial distribution into a target pile configuration. We use this task to evaluate hybrid modeling with a pretrained granular world model.

Data and Training Protocol. We collect whole-arm manipulation data for **Maze**, **Bathing**, and **Reach** using (i) RRT-generated feasible start/goal configurations and (ii) random-play trajectories obtained by applying noisy joint-velocity commands from feasible states. For **Granular**, we use the dataset released by [31]. All learning-based components (world models, value functions, and baselines) are trained *offline* on these datasets; no method uses online environment interaction for training.

Ablations (A1-4) and Baselines (B1-4). We perform all ablations in **Maze** to minimize confounds such as: **A1.** TACTIC-NoTacNoMask (Q1), which removes distributed tactile readings and the proximity mask, using RGB-D and proprioception only. **A2.** TACTIC-NoTac (Q1), which removes tactile readings f_t while retaining the vision-based proximity mask, isolating the effect of the mask when tactile information is unavailable. **A3.** TACTIC-NoCaSa (Q2), which replaces contact-aware sampling with the base STORM sampler [30]. **A4.** TACTIC-L / TACTIC-K / TACTIC (Q3), which evaluates learned latent dynamics only, kinematics only, or the full hybrid rollout model (Sec. IV-C).

Hybridization with pretrained world models. Given the general applicability of hybrid modeling, we evaluate it on manipulation tasks that extend beyond whole-arm manipulation. Specifically, we run experiments with pretrained world models: **B1.** V-JEPA2 [3] on Reach, and **B2.** DINO-WM [31] on Granular. In both cases, we add a kinematics-based goal cost to the planner to incorporate analytical structure on top of the learned predictor, while keeping the pretrained

Variant	SR $\uparrow$	TTC $\downarrow$	FV $\downarrow$
TACTIC-NoTacNoMask	63.5	20.3 $\pm$ 4.7	84.4 $\pm$ 53.9
TACTIC-NoTac	72.1	23.8 $\pm$ 8.5	55.1 $\pm$ 56.7
TACTIC	87.2	22.4 $\pm$ 8.1	39.0 $\pm$ 55.9

TABLE II: Contact-centric representation boosts task performance.

Method	FV $\downarrow$	Mean Jerk $\downarrow$
TACTIC-NoCaSa	111.9 $\pm$ 71.2	5.87 $\pm$ 2.98
TACTIC	39.0 $\pm$ 55.9	2.83 $\pm$ 3.23

TABLE III: Contact-aware sampling improves safety and smoothness.

encoder/predictor intact unless finetuning is explicitly stated.

Model-based baselines. We also compare TACTIC against two offline model-based RL baselines: **B3.** DreamerV3 [39] is a strong world-model RL method that learns a latent dynamics model together with actor/value functions. This provides a relevant comparison because it represents an alternative end-to-end way to act from a learned model—learned rollout and value-based policy—without explicit contact-aware sampling or hybrid analytical rollouts. **B4.** TD-MPC2 [40] is a competitive, decoder-free model-based method that learns latent dynamics and value estimates and can plan in latent space.

Metrics. We use the following metrics:

- **Success rate (SR):** % age of trials completing the task before timeout.
- **Time to completion (TTC):** time (in s) to success.
- **Force violations (FV):** count of unsafe-force timesteps

Results. Contact-centric sensing improves safety and performance. We ablate tactile sensing and proximity mask in **Maze**. Removing both degrades performance and increases unsafe interactions.

Contact-aware sampling reduces unsafe interactions and improves smoothness. Tab. III compares TACTIC against a variant that uses the base STORM sampler, keeping the predictive model and costs fixed. Injecting local contact geometry into the sampler yields substantially fewer force violations and smoother motion (lower jerk and action magnitude).

Hybrid rollouts improve success and reduce unsafe contact. Tab. IV compares latent-only, kinematics-only, and



Fig. 4: **Real-world task suite.** We evaluate across three real-world environments. The **3D Maze Environment** tests whole-arm navigation through static and dynamic obstacles, **Side Rollover** tests contact-rich manipulation to rotate a person while regulating interaction forces, and **Limb Repositioning** tests multi-contact support during lifting and repositioning of the legs to a target configuration. Whole-arm contact is highlighted in yellow.

Variant	SR $\uparrow$	TTC $\downarrow$	FV $\downarrow$
TACTIC-L	64.1	47.9 ± 13.4	108.3 ± 41.4
TACTIC-K	43.6	71.8 ± 4.3	79.9 ± 38.2
TACTIC	87.2	22.4 ± 8.1	39.0 ± 55.9

TABLE IV: TACTIC outperforms latent- and kinematic-only variants.

Method	Maze		Bed Bathing	
	SR $\uparrow$	FV $\downarrow$	SR $\uparrow$	FV $\downarrow$
TD-MPC2 [40]	65.1	93.9 ± 34.0	—	—
DreamerV3 [39]	75.9	82.4 ± 28.7	51.6	63.8 ± 17.4
TACTIC (ours)	87.2	39.0 ± 55.9	79.5	22.6 ± 45.2

TABLE V: TACTIC outperforms alternate model-based methods.

hybrid rollouts. Kinematics-only rollouts provide geometric grounding but cannot anticipate contact state changes. Latent-only rollouts capture interaction patterns but can drift without analytical structure. Hybrid rollouts achieve the best success/safety trade-off.

Hybrid modeling benefits pretrained world-model planning. We study two pretrained predictors: **V-JEPA2** on **Reach** and **DINO-WM** on **Granular**. In both cases, we compare the original planner to a *hybrid* variant that adds a kinematics-based goal cost during planning, keeping the pretrained encoder/predictor frozen unless finetuning is explicitly stated. Hybridization improves V-JEPA2 finetuning data efficiency (achieving comparable performance with $\sim 34\%$ less finetuning data) and improves geometric rollout quality for frozen DINO-WM (reducing Chamfer Distance by $\sim 26\%$; see Appendix I).

TACTIC outperforms alternate model-based approaches. Tab. V shows TACTIC achieves the highest success and lowest force violations on **Maze** compared to DreamerV3 and TD-MPC2, and substantially improves success and safety on **Bed Bathing** compared to offline DreamerV3.

VI. REAL WORLD EXPERIMENTS

While simulation validates our design choices in controlled settings, real-world deployment requires real-time control under sensing noise, actuation delays, and changing contact conditions. We evaluate TACTIC on a real robot in (i) a

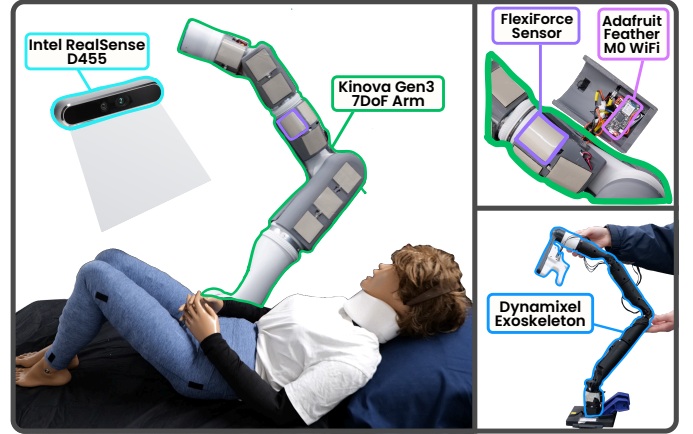


Fig. 5: **Real-world setup.** Kinova Gen3 7-DoF arm with 22 distributed tactile sensors, RGB-D perception, and exoskeleton-based teleoperation for data collection.

dynamic cluttered maze that tests contact-rich navigation involving highly constrained multi-contact trajectories, and (ii) contact-intensive manipulation tasks with a life-size manikin (~ 46 lbs, vinyl plastic) of the type used for nurse training on the tasks studied here, that require multi-contact regulation and goal-directed motion under safety limits. These experiments assess whether the contact-aware planning and hybrid modeling principles validated in simulation carry over to realistic manipulation scenarios.

Environments. We evaluate on 3 environments (Fig. 4):

- **3D maze environment.** Real-world instantiation of the maze task. Four central cylinders translate with amplitude 3 cm and speed 1 cm/s. The robot must reach a target pose while safely making contact when necessary.
- **Side rollover.** The robot rolls a manikin from a back-lying position onto its side while regulating contact forces to remain within safe limits.
- **Limb repositioning (LimbRepo).** The robot lifts and repositions the manikin’s legs to a target configuration (e.g., target joint angle or clinician-provided range-of-motion goal), often requiring multi-contact support across motion phases.

Robot Setup and Data Collection. We use a Kinova Gen3 7-DoF arm (Figure 5) with a novel tactile skin comprising

Task	Method	SR $\uparrow$	TTC $\downarrow$	FV $\downarrow$
Maze	Expert Teleop	5/5	30.4 $\pm$ 8.7	97.3 $\pm$ 50.1
	DP [42]	2/5	300 $\pm$ 45.2	0.0 $\pm$ 0.0
	TACTIC	3/5	71.4 $\pm$ 10.4	70.3 $\pm$ 29.4
Side Rollover	Expert Teleop	20/20	45.5 $\pm$ 13.6	141.2 $\pm$ 60.0
	DP [42]	3/20	275 $\pm$ 43.1	150.5 $\pm$ 32.0
	TACTIC	12/20	81.0 $\pm$ 19.8	104.3 $\pm$ 46.5
LimbRepo	Expert Teleop	20/20	30.4 $\pm$ 10.5	50.5 $\pm$ 24.0
	DP [42]	5/20	262 $\pm$ 53.0	61.4 $\pm$ 22.8
	TACTIC	14/20	74.4 $\pm$ 8.2	38.0 $\pm$ 16.5

TABLE VI: TACTIC outperforms DP on the Rollover and LimbRepo tasks. TACTIC also shows fewer force violations compared to the expert reference, i.e., the data it was trained on.

22 Tekscan FlexiForce sensors [41] mounted along the links. Sensor readings are transmitted through four Adafruit Feather M0 WiFi boards at ~ 800 Hz per board. RGB-D is captured by an overhead Intel RealSense D455. We collect (i) expert demonstrations via our custom exoskeleton-based teleoperation interface and (ii) random-play trajectories for coverage. We open-source the exoskeleton and WiFi-based distributed piezoresistive tactile sensing suite (see [website](#) for details). TACTIC is deployed as one layer of a multi-tier control stack. At the planning layer, an MPPI loop runs at ~ 12 Hz. Below the planner, a joint-space compliant controller running at 1 kHz tracks $\mathbf{u}_t^*$. Force thresholds, controller gains, the full set of safety-relevant constants are reported in Appendix C, and runtime statistics in Appendix A.

Baselines The model-based RL baselines (DreamerV3, TD-MPC2) underperformed TACTIC in our simulation benchmarks, and their real-world performance would be difficult to interpret. We compare against an offline imitation baseline and report expert teleoperation as a safety reference:

- **Diffusion Policy (DP) [42].** A widely used imitation learning baseline for real-world robot manipulation. It is appropriate in our setting because it can be trained directly from a limited number of expert demonstrations.
- **Human expert teleoperation reference.** We pair each autonomous trial with a teleop episode collected under the same environment configuration (start/goal parameters) and ask whether TACTIC can achieve comparable task performance while outperforming the reference in terms of force violations.

Results. Tab. VI shows TACTIC achieves higher success than Diffusion Policy on two out of three tasks while being comparable on Maze. DP shows no violation for maze since it usually enters from the top of the maze and fits into the final configuration without going “through” the maze. TACTIC produces fewer force-threshold violations than matched expert teleoperation episodes, indicating that contact-aware planning can regulate forces beyond demonstration distribution and do better than the human reference.

VII. DISCUSSION

We presented TACTIC, a receding-horizon controller framework for whole-arm manipulation that embraces whole-arm contact through contact-centric representation, contact-aware sampling, and hybrid modeling. Our results confirm that

jointly reasoning over analytical kinematics and learned latent dynamics outperforms pure learning approaches (DreamerV3, Diffusion Policy) in task performance and safety adherence. By shaping exploration via the contact Jacobian, TACTIC efficiently discovers valid sliding and pivoting motions. TACTIC does not provide formal safety guarantees. The low-level compliant controller makes the interactions comfortable under arbitrary planner output, while TACTIC reduces unsafe contact forces at the planning stage. Both complement safer hardware, watchdogs, fallbacks, and emergency stops in a broader deployment-ready stack. Future work can look into tighter integration with low-level control, where TACTIC also modulates stiffness and damping (variable impedance). Extending evaluation from manikins to human subjects is essential for validating both task performance and user comfort in caregiving settings. One scalability limitation of our framework is our reliance on a custom exoskeleton for high-fidelity data; we aim to reduce data dependence through cross-embodiment learning and investigate active tactile exploration to reduce state uncertainty.

ACKNOWLEDGEMENTS

We would like to thank Ben Dodson and Alexios Rekoutis for the initial design of the exoskeleton used for data collection. We would also like to thank Kapil Gangwar and Devesh Khilwani for help with debugging the circuit for the wireless sensor. This work was partly funded by National Science Foundation IIS #2132846, and CAREER #2238792. The Toyota Research Institute also partially supported this work. This article solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity.

REFERENCES

- [1] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, pages 2165–2183. PMLR, 2023.
- [2] Jose Barreiros, Andrew Beaulieu, Aditya Bhat, Rick Cory, Eric Cousineau, Hongkai Dai, Ching-Hsin Fang, Kunimatsu Hashimoto, Muhammad Zubair Irshad, Masha Itkina, et al. A careful examination of large behavior models for multitask dexterous manipulation. *arXiv preprint arXiv:2507.05331*, 2025.
- [3] Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zhohus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.
- [4] Quan Vuong, Sergey Levine, Homer Rich Walke, Karl Pertsch, Anikait Singh, Ria Doshi, Charles Xu, Jianlan Luo, Liam Tan, Dhruv Shah, et al. Open x-embodiment: Robotic learning datasets and rt-x models. In *Towards*

Generalist Robots: Learning Paradigms for Scalable Skill Acquisition@ CoRL2023, 2023.

- [5] Phillip M Grice, Marc D Killpack, Advait Jain, Sarvagya Vaish, Jeffrey Hawke, and Charles C Kemp. Whole-arm tactile sensing for beneficial and acceptable contact during robotic assistance. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pages 1–8. IEEE, 2013.
- [6] Advait Jain, Marc D Killpack, Aaron Edsinger, and Charles C Kemp. Reaching in clutter with whole-arm tactile sensing. *The International Journal of Robotics Research*, 32(4):458–482, 2013.
- [7] Marc D. Killpack and Charles C. Kemp. Fast reaching in clutter while regulating forces using model predictive control. In *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 146–153, 2013.
- [8] Kinova Robotics. Discover our gen3 robotic arm. <https://www.kinovarobotics.com/product/gen3-robots>. Accessed: 2026-05-21.
- [9] Boxin Xu, Luoyan Zhong, Grace Zhang, Xiaoyu Liang, Diego Virtue, Rishabh Madan, and Tapomayukh Bhattacharjee. Cushsense: Soft, stretchable, and comfortable tactile-sensing skin for physical human-robot interaction. *arXiv preprint arXiv:2405.03155*, 2024.
- [10] Zilin Si, Tianhong Catherine Yu, Katrene Morozov, James McCann, and Wenzhen Yuan. Robotswearer: Scalable, generalizable, and customizable machine-knitted tactile skins for robots, 2023.
- [11] Aimee Goncalves, Naveen Kuppaswamy, Andrew Beaulieu, Avinash Uttamchandani, Katherine M. Tsui, and Alex Alspach. Punyo-1: Soft tactile-sensing upper-body robot for large object manipulation and physical human interaction. In *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*, pages 844–851, 2022.
- [12] Zhe Liu, Diansheng Chen, Junlin Ma, Zhenyang Shen, Tianhao Wu, Zining Jia, and Yongkang Jiang. A warm hug from a robot: A dual-mode e-skin with programming compliance. *Review of Scientific Instruments*, 93(11), 2022.
- [13] Pascal Gliesche, Christian Kowalski, Max Pfingsthorn, and Andreas Hein. Geometry-based two-contact inverse kinematic solution for whole arm manipulation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8269–8274. IEEE, 2021.
- [14] Jose A Barreiros, Aykut Özgün Önel, Mengchao Zhang, Sam Creasey, Aimee Goncalves, Andrew Beaulieu, Aditya Bhat, Kate M Tsui, and Alex Alspach. Learning contact-rich whole-body manipulation with example-guided reinforcement learning. *Science Robotics*, 10(105):eads6790, 2025.
- [15] Rishabh Madan, Jiawei Lin, Mahika Goel, Angchen Xie, Xiaoyu Liang, Marcus Lee, Justin Guo, Pranav N Thakkar, Rohan Banerjee, Jose Barreiros, et al. Prioritouch: Adapting to user contact preferences for whole-arm physical human-robot interaction. *arXiv preprint arXiv:2509.18447*, 2025.
- [16] Alexis E Block, Sammy Christen, Roger Gassert, Otmar Hilliges, and Katherine J Kuchenbecker. The six hug commandments: Design and evaluation of a human-sized hugging robot with visual and haptic perception. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*, pages 380–388, 2021.
- [17] Michelle A. Lee, Yuke Zhu, Peter Zachares, Matthew Tan, Krishnan Srinivasan, Silvio Savarese, Li Fei-Fei, Animesh Garg, and Jeannette Bohg. Making sense of vision and touch: Learning multimodal representations for contact-rich tasks, 2019.
- [18] Roberto Calandra, Andrew Owens, Dinesh Jayaraman, Justin Lin, Wenzhen Yuan, Jitendra Malik, Edward H Adelson, and Sergey Levine. More than a feeling: Learning to grasp and regrasp using vision and touch. *IEEE Robotics and Automation Letters*, 3(4):3300–3307, 2018.
- [19] Yizhou Chen, Andrea Sipos, Mark Van der Merwe, and Nima Fazeli. Visuo-tactile transformers for manipulation. *arXiv preprint arXiv:2210.00121*, 2022.
- [20] Stephen Tian, Frederik Ebert, Dinesh Jayaraman, Mayur Mudigonda, Chelsea Finn, Roberto Calandra, and Sergey Levine. Manipulation by feel: Touch-based control with deep predictive models. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 818–824. IEEE, 2019.
- [21] Zhengtong Xu and Yu She. Letac-mpc: Learning model predictive control for tactile-reactive grasping. *IEEE Transactions on Robotics*, 2024.
- [22] Masashi Okada, Mayumi Komatsu, and Tadahiro Taniguchi. A contact model based on denoising diffusion to learn variable impedance control for contact-rich manipulation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7286–7293. IEEE, 2024.
- [23] Fangqi Zhu, Hongtao Wu, Song Guo, Yuxiao Liu, Chilam Cheang, and Tao Kong. Irasim: A fine-grained world model for robot manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9834–9844, 2025.
- [24] R Khorrambakht, Joaquim Ortiz-Haro, Joseph Amigo, Omar Mostafa, Daniel Dugas, Franziska Meier, and Ludovic Righetti. Worldplanner: Monte carlo tree search and mpc with action-conditioned visual world models. *arXiv preprint arXiv:2511.03077*, 2025.
- [25] Grady Williams, Andrew Aldrich, and Evangelos A Theodorou. Model predictive path integral control: From theory to parallel computation. *Journal of Guidance, Control, and Dynamics*, 40(2):344–357, 2017.
- [26] Rahima Khanam and Muhammad Hussain. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*, 2024.
- [27] Songrit Maneewongvatana and David M Mount. Analysis

- of approximate nearest neighbor searching with clustered point sets. *arXiv preprint cs/9901013*, 1999.
- [28] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.
 - [29] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, 31, 2018.
 - [30] Mohak Bhardwaj, Balakumar Sundaralingam, Arsalan Mousavian, Nathan D Ratliff, Dieter Fox, Fabio Ramos, and Byron Boots. Storm: An integrated framework for fast joint-space model-predictive control for reactive manipulation. In *Conference on Robot Learning*, pages 750–759. PMLR, 2022.
 - [31] Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. Dino-wm: World models on pre-trained visual features enable zero-shot planning. *arXiv preprint arXiv:2411.04983*, 2024.
 - [32] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
 - [33] Basile Terver, Tsung-Yen Yang, Jean Ponce, Adrien Bardes, and Yann LeCun. What drives success in physical planning with joint-embedding predictive world models? *arXiv preprint arXiv:2512.24497*, 2025.
 - [34] Ilya Kostrikov, Ashvin Nair, and Sergey Levine. Offline reinforcement learning with implicit q-learning. *arXiv preprint arXiv:2110.06169*, 2021.
 - [35] Ruolin Ye, Wenqiang Xu, Haoyuan Fu, Rajat Kumar Jenamani, Vy Nguyen, Cewu Lu, Katherine Dimitropoulou, and Tapomayukh Bhattacharjee. Rcareworld: A human-centric simulation world for caregiving robots. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 33–40. IEEE, 2022.
 - [36] Vikash Kumar, Rutav Shah, Gaoyue Zhou, Vincent Moens, Vittorio Caggiano, Abhishek Gupta, and Aravind Rajeswaran. Robohive: A unified framework for robot learning. *Advances in Neural Information Processing Systems*, 36:44323–44340, 2023.
 - [37] Lasse Sofus Møller Sørensen. Robohive Franka environment for V-JEPA2. <https://github.com/lassesofus/thesis>, 2026. Accessed: 2026-05-11.
 - [38] Kaifeng Zhang, Baoyu Li, Kris Hauser, and Yunzhu Li. Adaptigraph: Material-adaptive graph-based neural dynamics for robotic manipulation. *arXiv preprint arXiv:2407.07889*, 2024.
 - [39] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, April 2025.
 - [40] Nicklas Hansen, Hao Su, and Xiaolong Wang. Td-mpc2: Scalable, robust world models for continuous control, 2024.
 - [41] Tekscan, Inc. FlexiForce load/force sensors and systems. <https://www.tekscan.com/products/flexiforce-sensors>, 2024.
 - [42] Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.
 - [43] Marco Esposito. easy_handeye: Automated, hardware-independent hand-eye calibration for ROS. https://github.com/IFL-CAMP/easy_handeye. Accessed: 2026-07-09.

TABLE OF CONTENTS

A	System Overview and Latency Breakdown	13
1	Pipeline overview	13
2	Hyperparameter table	13
B	Contact-Centric Representation Details	13
1	Proximity mask construction	13
2	Encoder implementation details	13
C	Contact-Aware Action Sampling Details	13
1	Active contact set and contact geometry	13
2	Force weighting and projection operators	14
3	Safety projection QP	14
D	Hybrid Predictive Model and Planning Details	14
1	Two-fidelity rollouts for real-time control	14
E	Rollout Cost Details	15
1	Kinematic cost c_{kin}	15
2	Active force cost c_{active}	15
3	Predicted force cost c_{pred}	15
4	Latent goal cost c_{goal}	15
5	Weights and implementation	15
F	IQL Value Learning for Long-Horizon Tasks	15
1	Progress Variables and Subgoals	15
2	Latent Encoding and Goal Representation	16
3	Offline Transition Dataset	16
4	Goal-Conditioned IQL Critic	16
5	Using the Value for Terminal Shaping in MPPI	16
G	Tactile Skin: Hardware, Frames, Pose/Normal Calibration, and Limitations	16
1	Hardware overview	16
2	Sensor readings and preprocessing	16
3	Taxel pose, normals, and contact Jacobians	16
4	Force calibration	17
5	Driver and data integrity checks	17
6	Limitations	17
H	Data Collection Details	17
1	Exoskeleton teleoperation interface	17
2	Offline dataset summary	17
I	Experiment Details	17
1	V-JEPA2 and DINO-WM experiment details	17

A. SYSTEM OVERVIEW AND LATENCY BREAKDOWN

1. Pipeline overview

At each MPC step t , the system runs the following sequence: **Sensor inputs** $\rightarrow$ **Mask generation** $\rightarrow$ **Contact-centric observation** $\rightarrow$ **Encoding** $\rightarrow$ **Rollouts (draft + kinematic + full) + decoding** $\rightarrow$ **Scoring** $\rightarrow$ **Execution**. We acquire synchronized sensor data at 40 Hz. We then generate the proximity mask M_t (segmentation + kd-tree queries + taxel projection) which takes $\sim 5\text{--}8$ ms. We assemble the contact-centric observation o_t^{cc} and encode it into a latent $\mathbf{z}_t = \phi(o_t^{\text{cc}})$, updating the rolling context $\mathbf{z}_{t-L+1:t}$ (~ 6 ms). After receiving the sensor observation, in parallel, MPPI samples N action sequences and applies contact-aware shaping. For each candidate sequence, we roll out analytical kinematics $\mathbf{F}_k$ to obtain kinematic signals $\hat{s}_{t+1:t+H}^{(j)}$ and roll out the draft latent predictor $\mathbf{F}_t^{\text{draft}}$ for all particles to obtain predicted latents; these two rollouts are computed in parallel. We then select the top- M candidates by draft score and re-score only those with the full predictor, producing the final costs used to compute MPPI weights and update the nominal sequence. Finally, we execute the first m actions $\mathbf{u}_{t:t+m-1}^*$ using a low-level controller running at 1000 Hz (joint-space compliant controller), and replan at $t+m$ with updated observations. Overall, our MPC runs at close to 12 Hz.

2. Hyperparameter table

TABLE VII: Hyperparameters for TACTIC.

Parameter	Value
MPC horizon H	8
MPC time step Δt	0.025
MPPI particles N	100
Two-fidelity top- M	10
Latent history length L	3
Control mode	joint-space impedance

B. CONTACT-CENTRIC REPRESENTATION DETAILS

1. Proximity mask construction

We assume RGB and depth are spatially aligned and obtain camera intrinsics $\mathbf{K}$ using the `pyrealsense` API. We use a fixed extrinsic transform between the robot base and camera, which is needed only for projecting taxels into the image. We use the `easy_handeye` [43] package to obtain the extrinsics. For each pixel (u, v) with valid depth $z(u, v)$, we unproject to 3D in the camera frame as $\mathbf{p}(u, v) = z \mathbf{K}^{-1}[u, v, 1]^\top$. We form $P_r = \{\mathbf{p}(u, v) \mid (u, v) \in \Omega_r\}$ and $P_h = \{\mathbf{p}(u, v) \mid (u, v) \in \Omega_o\}$ from robot/object masks. The object masks are obtained using YOLO11n-seg. To train this model, we semi-automatically annotate the robot, maze, and manikin classes in our dataset using SAM2.1. After training, we convert the model to TensorRT and run the model with an inference time of 2–3 ms. We build a kd-tree on P_h and query each $\mathbf{p} \in P_r$ to compute $d(\mathbf{p}) = \min_{\mathbf{p}_h \in P_h} \|\mathbf{p} - \mathbf{p}_h\|_2$. We convert distances to an image-aligned proximity field by applying the following

linear clipping:

$$M_t^{\text{vis}}(u, v) = \text{clip}\left(1 - \frac{d(\mathbf{p}(u, v))}{d_{\text{max}}}, 0, 1\right),$$

where d_{max} is the maximum distance considered relevant for anticipating contact. Pixels outside Ω_r are set to zero. We render a separate blank mask M_t^{tax} (zeros, same resolution as the image). For each taxel k with measured force exceeding f_{min} , we compute its 3D pose via forward kinematics, transform it into the camera frame, project it with $\mathbf{K}$, and draw a filled disk of radius r pixels centered at the projected location. We fuse the visual proximity field and taxel indicator by per-pixel maximum and clip to $[0, 1]$:

$$M_t(u, v) = \max(M_t^{\text{vis}}(u, v), M_t^{\text{tax}}(u, v)).$$

In all experiments, we use $d_{\text{max}} = 0.05$ m, $f_{\text{min}} = 2$ N, and $r = 5$ px.

2. Encoder implementation details

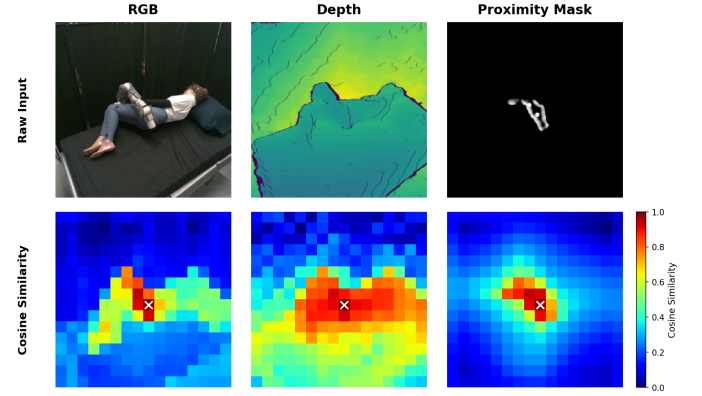


Fig. 6: Cosine similarity for RGB, Depth, and Proximity Mask.

Vision backbone. We apply the same DINOv3 backbone to I_t , D_t , and M_t in a batched forward pass. Depth and mask inputs are replicated to three channels and normalized using the same preprocessing as RGB (after scaling depth to metric units). The DINO feature map is pooled to a fixed-dimensional vector and projected to the vision embedding used by the world model. As shown in Fig. 6, the cosine similarity maps show that the encoder is able to extract spatially meaningful features from the inputs.

C. CONTACT-AWARE ACTION SAMPLING DETAILS

1. Active contact set and contact geometry

We define the active contact set $\mathcal{K}_t$ from taxels whose measured normal force exceeds threshold f_{min} . For each active taxel k , we treat the contact point as the taxel center obtained from forward kinematics. The corresponding translational Jacobian $\mathbf{J}_{k,t} \in \mathbb{R}^{3 \times n}$ is computed from the robot model. Each taxel has a known outward surface normal in link frame. We transform it to the world frame and negate it to obtain $\mathbf{n}_{k,t}$, so the normal points away from the contacted object and positive normal velocity $\mathbf{J}_{k,t}^n \mathbf{u} > 0$ corresponds to separation. This orientation is used by the safety projection (the shaping

operators are invariant to row sign) and by the compression update in Appendix E.

2. Force weighting and projection operators

We use the taxel force magnitude $f_{k,t}$ (scalar normal force) and define $\rho_{k,t} = \text{sat}(f_{k,t}/F_{\text{sat}})$ to reduce the influence of weak/noisy contacts while preventing a single contact from dominating via saturation. We compute $\mathbf{G}_t^{\text{force}}$ using the small matrix inverse $(\widetilde{\mathbf{J}}_t^{\text{n}} \widetilde{\mathbf{J}}_t^{\text{n}\top} + \varepsilon \mathbf{I})^{-1}$ with a stable factorization (e.g., Cholesky). When $K_t = 0$, we set $\mathbf{G}_t^{\text{force}} = \mathbf{0}$ and $\mathbf{G}_t^{\text{null}} = \mathbf{I}$. The ridge term $\varepsilon > 0$ prevents numerical issues when contact normals are redundant (e.g., nearly collinear rows in $\widetilde{\mathbf{J}}_t^{\text{n}}$).

Parameter	Value
σ_{force}	1.0
σ_{null}	0.7
σ_{min}	0.1
ε	10^{-4}

TABLE VIII: Hyperparameter values for contact-conditioned shaping and safety projection.

3. Safety projection QP

We form $\mathcal{K}_t^{\text{viol}} = \{k \in \mathcal{K}_t \mid f_{k,t} \geq F_{\text{safe}}\}$ and stack the corresponding rows into $\mathbf{J}_t^{\text{viol}}$. Because $|\mathcal{K}_t^{\text{viol}}|$ is small, we solve it efficiently with an active-set method. All 2^K active sets are enumerated, the corresponding equality-constrained least-squares problems are solved in closed form, and the feasible solution with minimum objective is selected. If $\mathcal{K}_t^{\text{viol}} = \emptyset$, we skip the projection.

Table VIII shows the parameters obtained in simulation by commanding the robot to regulate a fixed goal end-effector pose while injecting contacts of varying magnitudes and locations along the arm. We selected values that minimized safety violations and improved recovery behavior, using metrics including contact-limit violations, escape (separation) speed under high force, and tracking error.

The projection assumes contact normals are locally constant over the short MPC horizon, which is reasonable under our planning frequency. Since the constraint is applied only to contacts above F_{safe} and enforces separation rather than precise force control, errors in the assumed normals result in conservative behavior rather than unsafe actions.

Metric	Contact-aware	Contact-aware + QP
Violations	84.3 ± 39.8	50.0 ± 20.1

TABLE IX: Ablation results for QP-based projection.

We perform multi-contact ablation experiments to observe the effect of QP-based projection. Table IX reports the number of safety force violations under contact-aware control, comparing contact-conditioned shaping alone against shaping combined with our QP-based safety projection.

D. HYBRID PREDICTIVE MODEL AND PLANNING DETAILS

Each timestep is encoded into a latent vector $\mathbf{z}_t \in \mathbb{R}^{d_z}$ with $d_z = 384$. We form $\mathbf{z}_t$ by concatenation of modality embeddings from RGB, depth, proximity mask, forces, and proprioception. The predictor conditions on a context window $\mathbf{z}_{t-L+1:t}$ with $L = 3$. We implement $\mathbf{F}_l$ as a Transformer/Vision-Transformer-style predictor with 6 blocks, hidden size 384, and 8 attention heads. We use a frame-causal attention mask so that tokens from timestep t' attend only to timesteps $\leq t'$ within the context window. The predictor outputs $\hat{\mathbf{z}}_{t+1}$, and we obtain $\hat{\mathbf{z}}_{t+1:t+H}$ via iterative rollout.

Actions are commanded joint velocities $\mathbf{u}_t \in \mathbb{R}^n$. We normalize actions using per-dimension statistics (mean/std) and embed them with an MLP action encoder $\phi_{\text{act}} : \mathbb{R}^n \rightarrow \mathbb{R}^{d_a}$ with $d_a = 10$ and architecture 1D Conv (kernel=1, stride=1).

When predicting $\hat{\mathbf{z}}_{t+k}$, we condition on the step-aligned action $\mathbf{u}_{t+k-1}$ by injecting $\phi_{\text{act}}(\mathbf{u})$ into each Transformer block via AdaLN, i.e., the action embedding produces per-block scale/shift parameters.

Teacher-forced training uses the ground-truth context $\mathbf{z}_{t-L+1:t}$ to predict $\hat{\mathbf{z}}_{t+1:t+H}$. For the multistep rollout loss, we unroll autoregressively for k steps by updating the context with predicted latents using a sliding window: append $\hat{\mathbf{z}}$ and drop oldest. We use $K_{\text{roll}} = 2$ and weights $\lambda_k = 0.25$.

We attach lightweight decoder heads ψ_m to $\hat{\mathbf{z}}_{t+i}$ for planning-relevant signals. The binary contact decoder ψ_{contact} is a 2-layer MLP with hidden dims [64, 32] and ReLU activations that predicts per-taxel contact probabilities through a sigmoid output; we supervise it with a binary cross-entropy loss, with ground-truth contacts defined by thresholding measured forces at 2N. The joint-state decoder predicts $(\mathbf{q})$ (14D with sin/cos representation for $\mathbf{q}$) with an ℓ_2 loss. These decoded signals are used to derive force estimates for scoring (Appendix E). We train end-to-end (encoders + predictor + decoders) using AdamW with learning rate 5×10^{-4} (predictor), batch size $B = 32$, and 100 epochs. Segments are sampled as $(L + H)$ -length windows with random offsets.

1. Two-fidelity rollouts for real-time control

Draft model design and training. The draft predictor $\mathbf{F}_l^{\text{draft}}$ shares the same encoders and action embedding ϕ_{act} as the full predictor, but uses a lightweight two-layer MLP. We train $\mathbf{F}_l^{\text{draft}}$ purely using teacher forcing loss (imitate the full model).

Selection and rescoring. At each MPC step t , we sample N action sequences $\{\mathbf{u}_{t:t+H-1}^{(j)}\}_{j=1}^N$. We evaluate the analytical rollout $\mathbf{F}_k$ for all N particles. For the learned rollout, we: (i) roll out $\mathbf{F}_l^{\text{draft}}$ for all particles to obtain draft predictions and costs $\{\mathcal{J}_j^{\text{draft}}\}_{j=1}^N$, (ii) select the top- M candidates $\mathcal{I}_M = \text{TopM}(\{\mathcal{J}_j^{\text{draft}}\}, M)$ with $M = 10$ and $\mathcal{I}_M \subset \{1, \dots, N\}$, and (iii) re-roll out only those candidates with the full predictor $\mathbf{F}_l$ over the full horizon to compute $\{\mathcal{J}_j^{\text{full}}\}_{j \in \mathcal{I}_M}$.

Draft-error safeguard and fallback policy. To detect when the draft model is unreliable, we monitor a one-step latent

prediction error at runtime:

$$e_t = \|\hat{\mathbf{z}}_{t+1}^{\text{draft}} - \mathbf{z}_{t+1}\|_2,$$

where $\mathbf{z}_{t+1} = \phi(o_{t+1}^{\text{cc}})$ is computed from the next observation. If $e_t > \tau_{\text{draft}}$ with $\tau_{\text{draft}} = 0.1$, we disable the two-fidelity scheme and evaluate the full model for all particles (equivalently, set $M = N$) for the next $T_{\text{full}} = 4$ MPC steps.

E. ROLLOUT COST DETAILS

1. Kinematic cost c_{kin}

We decompose $c_{\text{kin}}(\hat{s}_{t+i})$ into feasibility and task terms:

$$c_{\text{kin}}(\hat{s}_{t+i}) = \lambda_{\text{lim}} c_{\text{lim}}(\hat{\mathbf{q}}_{t+i}) + \lambda_{\text{coll}} c_{\text{coll}}(\hat{\mathbf{q}}_{t+i}) + \lambda_{\text{ee}} c_{\text{ee}}(\hat{\mathbf{e}}_{t+i}) + \lambda_u c_u(\mathbf{u}_{t+i}), \quad (6)$$

with $\lambda_{\text{lim}} = 1000.0$, $\lambda_{\text{coll}} = 1000.0$, $\lambda_{\text{ee}} = [15.0, 10.0]$ (orientation, position), and $\lambda_u = 1.0$.

Joint limits. Let $(\mathbf{q}_{\min}, \mathbf{q}_{\max})$ be joint bounds and Δ_q a soft margin. We use

$$c_{\text{lim}}(\hat{\mathbf{q}}) = \sum_{m=1}^n \left([\hat{q}_m - (q_{\max,m} - \Delta_q)]_+^2 + [(q_{\min,m} + \Delta_q) - \hat{q}_m]_+^2 \right), \quad (7)$$

with $\Delta_q = 0.05 (q_{\max,m} - q_{\min,m})/2$.

Self-collision. We follow STORM's implementation and learn a lightweight collision classifier for our robot model.

Task-space tracking (if specified). When a task-space end-effector target pose $\mathbf{e}_g = (\mathbf{p}_g, \mathbf{R}_g)$ is defined, we penalize translation and rotation:

$$c_{\text{ee}}(\hat{\mathbf{e}}) = \frac{\|\hat{\mathbf{p}} - \mathbf{p}_g\|_2^2}{\sigma_p^2} + \frac{\|\mathbf{I} - \mathbf{R}_g^\top \hat{\mathbf{R}}\|_F^2}{\sigma_R^2}. \quad (8)$$

2. Active force cost c_{active}

Let $\mathcal{K}_t$ denote the active contact set at time t . For each $k \in \mathcal{K}_t$, we form a spring-based force estimate along the measured contact normal using the kinematic rollout $\hat{\mathbf{q}}_{t+1:t+H}$. For taxel k , let $\mathbf{J}_{k,t}^n = \mathbf{n}_{k,t}^\top \mathbf{J}_{k,t}$ be the normal Jacobian row. Given a kinematic rollout $\hat{\mathbf{q}}_{t:t+H}$, define $\Delta \mathbf{q}_{t+i} = \hat{\mathbf{q}}_{t+i} - \hat{\mathbf{q}}_{t+i-1}$ and the predicted normal displacement

$$\Delta x_{k,t+i}^n = \mathbf{J}_{k,t+i}^n \Delta \mathbf{q}_{t+i}. \quad (9)$$

We treat motion into contact as compression and accumulate

$$\hat{\delta}_{k,t+i} = \max(0, \hat{\delta}_{k,t+i-1} - \Delta x_{k,t+i}^n), \quad \hat{\delta}_{k,t} = 0, \quad (10)$$

then estimate normal force with a linear spring

$$\hat{F}_{k,t+i} = k_{\text{spr}} \hat{\delta}_{k,t+i}, \quad (11)$$

where k_{spr} is fit offline. For active contacts, we initialize the spring compression using the measured force at time t to reduce bias in the rollout. For each $k \in \mathcal{K}_t$ with measured normal force $f_{k,t}$, we set

$$\hat{\delta}_{k,t} = \text{clip}\left(\frac{f_{k,t}}{k_{\text{spr}}}, 0, \delta_{\max}\right), \quad (12)$$

and then propagate $\hat{\delta}_{k,t+i}$ using Eq. (10). The resulting force estimate is $\hat{F}_{k,t+i} = k_{\text{spr}} \hat{\delta}_{k,t+i}$. We penalize violations of the safety threshold F_{safe} over the active contacts with

$$c_{\text{active}}(\hat{\mathbf{q}}_{t+i}, \mathbf{f}_t) = \frac{1}{|\mathcal{K}_t|} \sum_{k \in \mathcal{K}_t} [\hat{F}_{k,t+i} - F_{\text{safe}}]_+^2. \quad (13)$$

3. Predicted force cost c_{pred}

Instead of decoding raw forces directly from the predicted latents (which can be noisy when evaluated far into the horizon), we derive a stable force signal for scoring from the decoded binary contacts and joint positions. We first use the binary contact decoder ψ_{contact} to estimate the contact onset time $\hat{\tau}_k$ for each taxel k . We record the predicted joint configuration at onset, $\hat{\mathbf{q}}_{t+\hat{\tau}_k}$, obtained from the joint-state decoder, and treat it as the reference configuration for subsequent force estimation. We then apply the same spring model (with stiffness k_{spr} as above) to convert predicted compression after onset into a per-taxel normal force estimate $\hat{\mathbf{f}}_{t+i} \in \mathbb{R}^{N_s}$.

We penalize predicted forces above the safety threshold F_{safe} with a hinge-squared penalty:

$$c_{\text{pred}}(\hat{\mathbf{f}}_{t+i}) = \frac{1}{N_s} \sum_{k=1}^{N_s} [\hat{\mathbf{f}}_{k,t+i} - F_{\text{safe}}]_+^2. \quad (14)$$

The force safety cost c_{force} used in the main paper combines the active and predicted terms:

$$c_{\text{force}}(\hat{s}_{t+i}, \hat{\mathbf{z}}_{t+i}, \mathbf{f}_t) = c_{\text{active}}(\hat{\mathbf{q}}_{t+i}, \mathbf{f}_t) + c_{\text{pred}}(\hat{\mathbf{f}}_{t+i}). \quad (15)$$

4. Latent goal cost c_{goal}

We encode the goal observation as $\mathbf{z}_g = \phi(o_g^{\text{cc}})$, where o_g^{cc} is formed analogously to o_t^{cc} (Sec. IV-A). The goal cost is

$$c_{\text{goal}}(\hat{\mathbf{z}}_{t+i}, \mathbf{z}_g) = \|\hat{\mathbf{z}}_{t+i} - \mathbf{z}_g\|_2^2. \quad (16)$$

5. Weights and implementation

Costs are summed over the horizon with discount factor $\gamma = 0.98$. Infeasible rollouts (large joint-limit violations or self-collision) are handled by increasing the corresponding penalty by a factor of 1000 rather than rejecting the sample.

F. IQL VALUE LEARNING FOR LONG-HORIZON TASKS

1. Progress Variables and Subgoals

For each long-horizon task, the dataset provides low-dimensional progress variables $\xi_t \in \mathbb{R}^{d_g}$ (e.g., body turn angle, limb joint angles). We obtain the values for these indicators during data collection by mounting ArUco markers on the head and limb joints. Note that these markers are not required at test time. Instead, we train progress heads that predict these progress variables for the current $\mathbf{z}_t$. At test time we specify an ordered subgoal sequence $\{\xi_g^{(1)}, \dots, \xi_g^{(G)}\}$ and maintain an active index g during execution. We advance to the next subgoal when the measured progress satisfies

$$\|\xi_t - \xi_g^{(g)}\|_2 \leq \delta_\xi, \quad (17)$$

for a fixed threshold δ_ξ (task-dependent).

2. Latent Encoding and Goal Representation

We freeze the multimodal encoders (Sec. IV-A) and compute a latent embedding $\mathbf{z}_t$ for each observation. For each subgoal $\xi_g^{(g)}$, we also form a corresponding goal observation and embed it with the same frozen encoders to obtain the goal latent $\mathbf{z}_g$. The choice of subgoals depends on task difficulty. For instance, in case of rollover, the subgoals are chosen as the two intermediate frames where contact switching happens (robot pulls from near the knee and then moves closer to the lower back and pulls from there).

3. Offline Transition Dataset

From the expert dataset, we construct an offline transition set

$$\mathcal{D} = \{(\mathbf{z}_t, \mathbf{u}_t, \mathbf{z}_{t+1}, \mathbf{z}_g, r_t, d_t)\}, \quad (18)$$

where $\mathbf{u}_t$ is the executed action, $d_t \in \{0, 1\}$ indicates episode termination, and the reward is defined as the negative of the cost used for scoring task state (Sec. IV-D):

$$r_t = -\ell_t. \quad (19)$$

ℓ_t includes a progress/subgoal tracking term built from the labeled ξ_t and current subgoal $\xi_g^{(g)}$, e.g.,

$$\ell_{\text{prog}}(\xi_t, \xi_g^{(g)}) = \|\xi_t - \xi_g^{(g)}\|_2^2. \quad (20)$$

For rollover, this is the L2 error between the goal head angle and current head angle (predicted by the progress head). In case of limb repositioning, this is the L2 error between goal limb configuration and current limb configuration.

4. Goal-Conditioned IQL Critic

We implement goal-conditioning by concatenating the state and goal latents (and optionally their difference):

$$\mathbf{x}_t = [\mathbf{z}_t, \mathbf{z}_g, (\mathbf{z}_g - \mathbf{z}_t)]. \quad (21)$$

We learn double action-value functions $Q_{\phi_1}(x_t, \mathbf{u}_t)$, $Q_{\phi_2}(x_t, \mathbf{u}_t)$ and a value function $V_\theta(x_t)$, with a slowly-updated target value network $\bar{V}_\theta$.

Q-learning objective. The TD target is

$$y_t = r_t + \gamma(1 - d_t) \bar{V}_\theta(x_{t+1}), \quad (22)$$

and we minimize

$$\mathcal{L}_Q = \mathbb{E}_{\mathcal{D}} \left[(Q_{\phi_1}(x_t, \mathbf{u}_t) - y_t)^2 + (Q_{\phi_2}(x_t, \mathbf{u}_t) - y_t)^2 \right]. \quad (23)$$

Expectile value objective. Following IQL, V_θ is trained via expectile regression toward the in-dataset action values:

$$\mathcal{L}_V = \mathbb{E}_{\mathcal{D}} [\rho_\tau(\min(Q_{\phi_1}, Q_{\phi_2})(x_t, \mathbf{u}_t) - V_\theta(x_t))], \quad (24)$$

where ρ_τ is the expectile loss with expectile parameter $\tau \in (0, 1)$.

We use discount $\gamma = 0.99$, expectile $\tau = 0.7$, and Polyak target update rate $\alpha = 0.005$. All networks (Q_{ϕ_1} , Q_{ϕ_2} , V_θ , and the AWR actor π_ψ) are 2-layer MLPs, dropout $p = 0.1$, and ReLU activations. We train all networks with AdamW (lr = 3×10^{-4}). The advantage-weighted regression temperature is $\beta_{\text{AWR}} = 3.0$.

5. Using the Value for Terminal Shaping in MPPI

At test time, MPPI rolls out the learned latent predictor $\mathbf{F}_l$ to obtain the horizon-end latent $\hat{\mathbf{z}}_{t+H}$ for each sampled action sequence. We then augment the rollout cost as

$$\mathcal{J}_{\text{aug}}(\mathbf{u}_{t:t+H-1}) = \mathcal{J}(\mathbf{u}_{t:t+H-1}) - \lambda_V V_\theta(\hat{\mathbf{z}}_{t+H}, \mathbf{z}_g), \quad (25)$$

where V_θ is the learned IQL value function. Since V_θ estimates reward-to-go and MPPI minimizes cost, the negative sign converts reward-to-go into a cost-to-go term. We set $\lambda_V = 0.1$.

Method	SR
TACTIC	6/10
TACTIC (w/o Value Fn.)	2/10

TABLE X: Ablation on the effect of the terminal value function, reported as success rate (SR) over 10 trials.

Ablation with value function. We ablate the terminal value function in TACTIC to assess its contribution to task success in maze environment, reporting success rate (SR) over 10 trials.

G. TACTILE SKIN: HARDWARE, FRAMES, POSE/NORMAL CALIBRATION, AND LIMITATIONS

1. Hardware overview

We mount 22 Tekscan FlexiForce A502 force-sensing resistors (FSRs) on the robot to measure distributed normal contact forces along the arm. Each pad measures normal force over a $50.8 \text{ mm} \times 50.8 \text{ mm}$ sensing area and outputs a resistance change proportional to applied load. Sensors are grouped into four physical mounts: the two wrist joints (5 pads each), the forearm link (6 pads), and the upper-arm link (6 pads). Each mount is instrumented with an Adafruit Feather M0 WiFi board that streams raw sensor readings over UDP at 800 Hz (per board).

2. Sensor readings and preprocessing

Taring. Before interaction, we tare each sensor by estimating its no-contact offset from a recent window of raw readings. A tare is accepted only when the window variance is below a stability threshold.

Filtering. We apply a median filter followed by a low-pass filter to the tared signal. The resulting tactile vector $\mathbf{f}_t \in \mathbb{R}^{N_s}$ is published at 100 Hz for control and logging.

3. Taxel pose, normals, and contact Jacobians

Taxel extrinsics. For each taxel k , we define a taxel frame $\{T_k\}$ attached to the sensing pad and store its fixed transform with respect to the corresponding link frame $\{L\}$, ${}^L\mathbf{T}_{T_k}$. We obtain taxel positions and outward normals in the link frame from manual measurements (vernier caliper) and CAD alignment, and include these extrinsics in the robot model.

Runtime calculation. At runtime, forward kinematics provides the link pose ${}^W\mathbf{T}_L(\mathbf{q}_t)$, which we use to transform each taxel pose and normal into the world frame and to compute the translational Jacobian at the taxel center. We validate these transforms by visualizing taxel frames across diverse arm configurations.

4. Force calibration

We calibrate the skin sensor using an ATI Nano25 force/torque sensor. We repeatedly apply normal loads over a range of magnitudes and fit a linear model that maps raw readings to force in Newtons using least-squares regression.

5. Driver and data integrity checks

We implement a `skin_driver` that receives UDP packets from the WiFi boards and exposes ROS services and topics for taring and reading tactile values. The driver includes a watchdog for common failure modes (e.g., missing packets, low battery, disconnected boards, inactive channels). The driver performs the filtering discussed above and publishes the filtered forces as an array.

6. Limitations

Coverage. The tactile skin provides discrete coverage over a subset of the arm surface; contact can occur in uninstrumented regions. **Normal-only sensing.** The sensors measure normal force only. This ignores tangential forces and frictional effects (e.g., shear during sliding), which can be informative for both scoring and sampling in contact-rich tasks. **System dependence.** Our method benefits from improved tactile coverage, higher bandwidth, and better calibration. The core algorithm does not assume a specific sensor type, but performance will improve depending on the quality of contact localization and force estimation available at runtime.

H. DATA COLLECTION DETAILS

1. Exoskeleton teleoperation interface

We collect human demonstration data using a custom exoskeleton teleoperation interface that enables mirrored whole-arm control of a Kinova Gen3 (7-DoF). Mirroring human arm motion is important for capturing whole-arm contact strategies relevant to caregiving, including distributing contact across links and coordinating intermediate-link and end-effector motion. The exoskeleton is a scaled replica of the robot and matches its kinematic structure using seven Dynamixel XL330-M288-T actuators. An additional Dynamixel XL330-M077-T actuator in a trigger handle is used to start and stop data collection. We include safety checks and a paired initialization procedure to ensure consistent alignment between the exoskeleton and robot before teleoperation.

2. Offline dataset summary

During data acquisition, RGB-D, tactile, and proprioceptive observations are recorded synchronously at 40 Hz. RGB and depth images are captured with an Intel RealSense D455 at 60 Hz and time-synchronized to the 40 Hz logging stream. In the 3D Maze environment, we collected 125 expert and 125 random-play episodes (each ~ 30 s). For Side Rollover, we collected 100 expert and 100 random-play episodes. For Limb Repositioning, we collected 100 expert and 100 random-play episodes. Episodes in the bed-based environments have an average duration of ~ 45 s. Expert trajectories correspond to successful task completion (e.g., reaching the specified

goals in the maze), while random-play trajectories consist of exploratory motions that still progress toward task-relevant goal states.

I. EXPERIMENT DETAILS

1. V-JEPA2 and DINO-WM experiment details

Finetuning V-JEPA2 and Hybrid Model Experiment. We load the pre-trained encoders and perform action-conditioned finetuning of the predictor using the tabletop reach dataset. To make the model hybrid, we do FK to compute end-effector samples and add a cost for end-effector goal reaching. This cost forces the planner to blindly move along the line connecting the current state and goal state, thus not reasoning about collision avoidance at all. We combine this with the latent goal reaching cost which intuitively encodes the avoidance behavior seen in the dataset. Upon tuning the weights for 5 different start-goal configurations (not included in testing) we fix the cost weights and run evaluations with models that are trained on different percentages of finetuning data. We also run the experiment with different obstacle shapes and sizes, and obtain similar performance improvements since the kinematics-based policy is independent of these effects in the data.