Opt2VLA: Force-Aware Vision-Language-Action for Contact-Rich Humanoid Whole-Body Manipulation

Fukang Liu*, Yipu Chen*, Jaehwi Jang, Danfei Xu, Zsolt Kira, and Ye Zhao

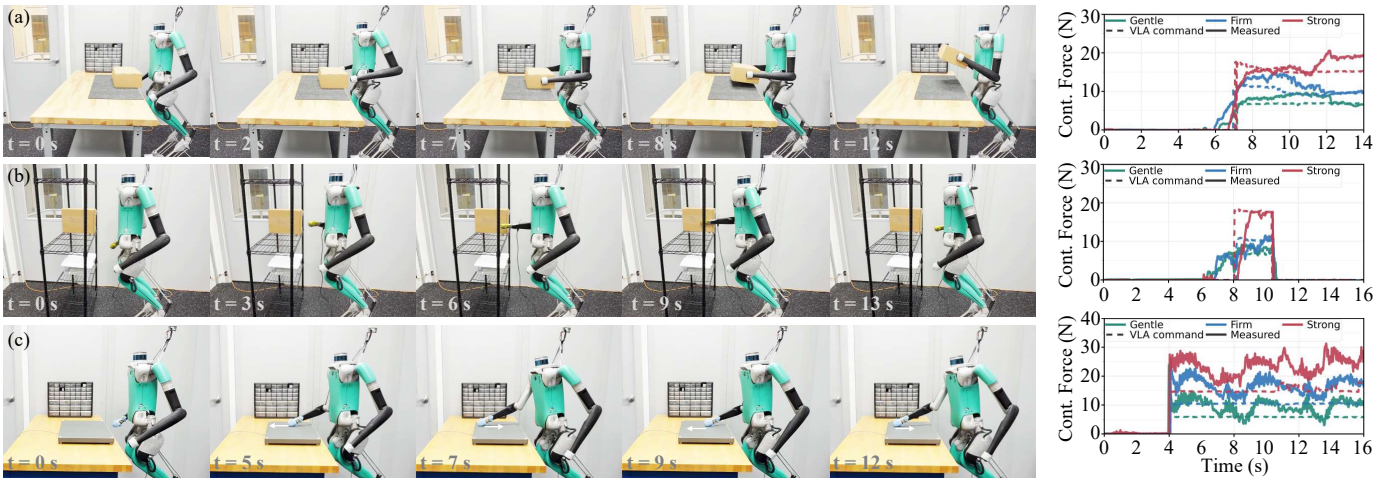

Fig. 1. **Opt2VLA enables autonomous, contact-rich whole-body humanoid manipulation with language-specified force levels.** (a) box pickup, (b) shelf-box pushing against a designated shelf wall, and (c) surface wiping. Snapshots show the task progression. Plots compare VLA-issued force commands (dashed) and measured contact forces (solid) for selected hardware trials under *gentle*, *firm*, and *strong* force prompts. Opt2VLA jointly predicts motion and contact-force commands, enabling language-conditioned force modulation during autonomous whole-body manipulation.

***Abstract*—Humanoid robots are expected to perform diverse human-level tasks in daily environments, many of which require precise regulation of interaction forces. While recent vision-language-action (VLA) models have shown promise for semantic planning and visuomotor control, existing humanoid systems primarily represent actions through geometric motion goals and rely on whole-body controllers focused on motion tracking, with limited explicit reasoning or control of interaction forces. This limitation is particularly relevant in contact-rich tasks, where geometrically similar motions may require different force regimes depending on the task context and where visual observations may become unreliable after contact. In this work, we present Opt2VLA, a force-aware VLA framework that introduces explicit force commands at the VLA-to-control interface for humanoid whole-body manipulation. A single multi-task VLA policy jointly predicts both geometric motion goals and continuous contact-force references, which are tracked by task-specific reinforcement learning (RL)-based whole-body controllers. To provide scalable and physically grounded supervision, we generate dynamically feasible and contact-consistent training data via whole-body trajectory optimization (TO) with explicit force references. We evaluate Opt2VLA on three contact-rich humanoid tasks and show that explicit force conditioning enables more accurate and consistent force regulation than motion-only control, while physically grounded torque supervision from TO further improves force tracking accuracy and stability. Closed-loop evaluations further demonstrate language-conditioned force modulation with Opt2VLA in simulation and on humanoid hardware.** https://opt2vla.github.io

I. Introduction

Humanoid robots are designed to operate in human-centered environments, where many everyday tasks require sustained physical interaction with the world [1], [2]. Beyond geometric accuracy, such interactions demand precise regulation of contact forces to ensure safety, stability, and task effectiveness, for example when pushing objects into constrained spaces, maintaining contact while sliding along surfaces, or interacting under partial visual occlusion. As humanoid systems move beyond scripted motions toward autonomous operation, the ability to reason about and control interaction forces becomes increasingly central to robust whole-body behavior.

Recent progress in whole-body control (WBC) has enabled robust humanoid behaviors across a wide range of locomotion and manipulation tasks, demonstrating impressive performance in the real world [2]–[4]. These advances provide a strong foundation for humanoid low-level execution. In parallel, vision-language-action (VLA) models have shown promise in enabling robots to interpret natural language instructions and plan long-horizon behaviors from visual observations [5]–[8]. Recent humanoid systems have begun integrating these capabilities with whole-body control, showing progress toward instruction-driven whole-body behaviors [7], [9]–[12]. However, these approaches primarily use motion-based action representations, such as end-effector poses or joint-position targets, with limited emphasis on explicitly specifying desired interaction forces at the VLA-to-control interface.

Despite the growing success of humanoid VLA models, this motion-based abstraction can be insufficient for contact-rich tasks. Geometric motion goals alone may not fully specify the intended interaction, since similar motions may require different levels of contact forces depending on the task context, object properties, or safety constraints. Recent contact-aware and tactile-based VLA methods have begun incorporating force

* Equal contribution. Authors are with the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA, USA.

or tactile feedback into manipulation or loco-manipulation [13]–[15]. However, these approaches generally use contact signals as observations or feedback for action prediction, rather than explicitly specifying desired interaction force as a high-level reference jointly planned with motion. Incorporating force into a humanoid autonomy pipeline, however, presents additional challenges. At the high-level, force references should remain compact enough for VLA prediction while providing sufficient information for robust whole-body execution under uncertain contact conditions. Learning to predict such references also requires structured supervision that aligns motion goals and force references with visual observations and language instructions. Obtaining this supervision is particularly challenging for humanoids, for which collecting contact-rich demonstrations through teleoperation requires substantial effort [7], [10], and providing diverse and consistently labeled force references adds further requirements to the data collection process.

These observations motivate an explicit motion-force interface in humanoid VLA systems. We separate task-level specification of the of the desired interaction force from its joint-level realization by whole-body control. Although force can also be modulated indirectly through motion commands [16], we expose the desired contact force explicitly alongside geometric motion goals, so that the high-level policy can specify both where to move and how strongly to interact.

In this work, we present **Opt2VLA**, a force-aware VLA framework for humanoid whole-body control, as shown in Fig. 2. Opt2VLA is built around two complementary ideas. First, a single multi-task VLA policy, instantiated from a pretrained VLA model [7], jointly predicts geometric motion goals and continuous contact-force references from task instructions, visual observations, robot states, and measured joint-torque feedback. Second, we use task-specific force-conditioned control with torque supervision (FCT), where TO-derived joint-torque references provide training-time guidance to improve force realization. Rollouts of these controllers further provide multimodal data paired with motion, force, and language targets for VLA fine-tuning, without requiring large-scale teleoperated data collection. This hierarchy decouples task-level motion-force prediction from low-level execution while allowing the predicted force commands to directly condition whole-body control.

We evaluate the proposed framework on a set of contact-rich humanoid tasks that require precise force regulation. The results show improved force regulation over the evaluated motion-tracking baseline, while physically grounded torque supervision from TO further enhances force tracking accuracy and stability. Our contributions are:

- We formulate force-aware VLA control for humanoid whole-body systems through an explicit interface of motion and contact-force commands.
- We present a hierarchical framework that integrates vision-language planning with force-aware RL-based whole-body control, using TO-generated references for controller training and controller rollouts for VLA fine-tuning.
- We demonstrate improved force regulation through explicit force conditioning and TO-derived torque supervision, and validate the framework through sim-to-real transfer on contact-rich humanoid tasks. We will publicly release the force-aware humanoid dataset collected through rollouts of controllers trained using TO-generated references.

II. Related Work

A. Humanoid Whole-Body Control

Model-based control and optimization methods that rely on accurate dynamic models have long served as a foundation for humanoid whole-body behaviors [3], [17]–[20]. However, their dependence on precise model fidelity and contact assumptions makes real-time deployment for complex humanoid interactions challenging under uncertainty in robot dynamics and environment contacts. Whole-body force regulation has also been demonstrated on position-controlled humanoids [16].

RL-based approaches, including motion imitation, have emerged as prominent methods for humanoid whole-body control, leveraging large-scale simulation and task-specific objectives to acquire diverse behaviors [21]–[23]. A common paradigm guides humanoid motion using expert trajectories retargeted from human demonstrations [4], [24]–[26]. While such pipelines have enabled versatile motions on humanoid hardware, the embodiment gap between humans and humanoids makes the retargeting challenging, and the resulting trajectories are not always kinematically or dynamically feasible, which can degrade data quality. Moreover, the absence of explicit force and torque supervision in these motion priors limits their ability to provide physically grounded supervision for contact-rich interactions requiring precise force regulation. Relatedly, SoftMimic [27] augments motion imitation with compliant reference trajectories to enable stiffness-modulated whole-body behaviors under external disturbances, but does not explicitly represent or command task-level interaction forces. To address these challenges, recent work has explored combining model-based optimization with RL for legged-robot control [28]–[31]. TO provides dynamically feasible references by incorporating robot dynamics and contact constraints, and has been increasingly leveraged to guide learning-based locomotion and whole-body control [32]–[36]. More recently, Opt2Skill [37] leverages full-order dynamic TO for humanoid loco-manipulation, showing that dynamically feasible motion and torque references improve motion tracking and contact-rich execution. However, these methods are primarily conditioned on geometric motion references and do not expose desired contact force as an independent task-level command. As a result, different interaction-force requirements cannot be specified directly from high-level task instructions without modifying the motion reference or controller.

B. Vision-Language-Action for Humanoids

Recent advances in VLA models demonstrate that large-scale multimodal pretraining can endow robots with strong generalization and language grounding [5], [6], motivating their extension to humanoid autonomy [38]. Recent humanoid-focused VLA systems ground language and perception into whole-body actions, typically represented as kinematic targets such as end-effector poses or joint configurations. Existing

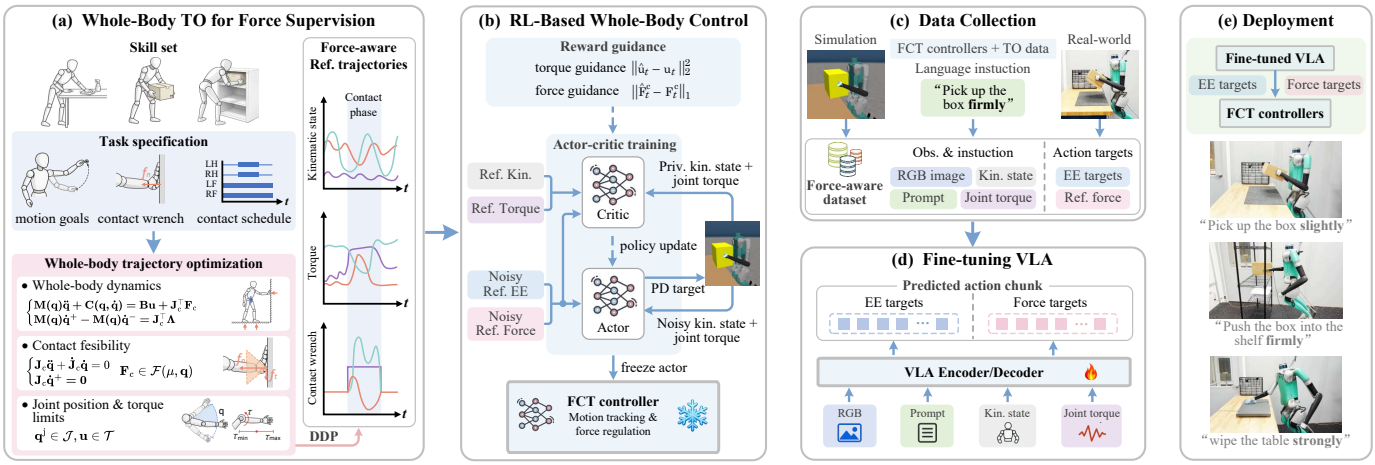

Fig. 2. **Overview of Opt2VLA**. (a) Whole-body TO generates physically consistent motion, joint torque, and contact force trajectories under full-body dynamics and contact constraints. (b) Force-conditioned controller with torque supervision (FCT) is trained to track geometric motion goals (i.e., end-effector 3D target positions) while regulating interaction force, using privileged torque supervision only during training. (c) Large-scale simulation rollouts, together with a smaller set of real-world executions, are used to collect multimodal data including vision, language prompts, robot kinematic state, measured joint torques, end effector targets, and target-force labels. (d) A pretrained vision–language–action model is then fine-tuned to predict motion goals and contact-force references from language, vision, kinematic state, and joint torque, and (e) deployed to execute force-aware behaviors through the FCT controllers.

approaches use hierarchical planning or latent action representations to enable locomotion, manipulation, and long-horizon loco-manipulation behaviors on real humanoids [9]–[11], [39]. Recent open VLA models trained on diverse robot datasets expand the scope of generalized humanoid skills [7].

Despite these advances, existing humanoid VLA formulations largely abstract away physical interaction forces during planning, leaving them to downstream controllers rather than explicitly specifying the desired interaction regime. Recent studies on tabletop manipulation [13], [14] incorporate joint-torque or end-effector force feedback into pretrained VLA models and shows improved performance on contact-rich tasks. FWBC-VLA [15] estimates interaction strength from joint-torque residuals and uses this feedback to condition VLA action generation and whole-body compensation on a wheeled-legged platform. Together, these studies demonstrate the value of physical interaction feedback, but such feedback describes the ongoing interaction rather than specifying the force required by the task. In contrast, our work introduces explicit force commands at the interface between humanoid vision-language planning and whole-body control, enabling desired interaction forces to be predicted from task context and directly regulated during execution.

III. Method

We consider contact-rich humanoid manipulation tasks specified by natural language instructions and visual observations, where successful execution requires both geometric motion and appropriate interaction force. Our framework adopts a hierarchical design that separates high-level planning from low-level execution. A VLA policy, instantiated from a pretrained VLA model, predicts geometric motion goals together with contact-force references specifying the desired interaction. An RL-based whole-body controller executes these commands by jointly tracking motion and regulating interaction force. To provide scalable and physically grounded supervision without teleoperation, we use whole-body TO to generate dynamically feasible motion, contact-force, and joint-torque references for controller training. Rollouts of the resulting controllers then provide multimodal training data for VLA fine-tuning.

A. VLA Policy with Explicit Force Commands

Our Opt2VLA generates high-level motion and force commands for whole-body execution Given a language instruction l, visual observations I_t, and histories of robot states $\mathbf{s}_{t-K:t}$ and measured joint torques $\boldsymbol{\tau}_{t-K:t}$, the policy predicts a sequence of geometric motion goals together with temporally aligned contact-force references:

$$\left(\hat{\mathbf{P}}^{\mathrm{ee}}_{t:t+H}, \hat{\mathbf{F}}^{\mathrm{ref}}_{t:t+H}\right) = \pi_{\mathrm{Opt2VLA}}\left(l, I_t, \mathbf{s}_{t-K:t}, \boldsymbol{\tau}_{t-K:t}\right), \quad (1)$$

where $\hat{\mathbf{P}}^{\mathrm{ee}}_{t:t+H}$ denotes the predicted 3D position waypoints of the left and right hands and feet over horizon H, and $\hat{\mathbf{F}}^{\mathrm{ref}}_{t:t+H}$ denotes the corresponding normal-contact-force sequences for the active manipulation contact. We instantiate the VLA policy from a pretrained GR00T-N1.7 model [7] and augment its input space with upper-body sensed joint torque history and action space with contact-force references, enabling joint prediction of geometric motion and force commands.

Torque Conditioning. We investigate how physical interaction feedback supports the prediction of motion and contact-force references. Inspired by [14], we compare incorporating measured joint torques via the native GR00T state encoder against using a dedicated torque encoder that feeds an additional token into the action decoder. We also evaluate whether temporal state and torque histories improves closed-loop execution. For selected variants, we further augment the policy with an auxiliary head that predicts future measured joint torques from the same rollout sequence. These predictions are supervised using recorded robot torques and are used only as an auxiliary learning objective; the VLA control outputs remain the predicted motion goals and contact-force references. We evaluate these choices through closed-loop ablations measuring task completion rate, agreement with language-specified force levels, and force-tracking accuracy.

Data Collection and Language Annotation. Training data is collected in simulation by rolling out TO-generated motions using the force-aware controller under different target force conditions. Across trajectories, we vary task configurations such as target force, object pose, and object properties while keeping the corresponding motion reference and force command synchronized. During rollout, we record egocentric RGB observations, robot proprioceptive states, measured joint torques, geometric motion targets, and the associated continuous target-force labels.

Each trajectory is paired with a natural-language instruction describing the task and desired interaction regime. We use the descriptors "*gently*", "*firmly*", and "*strongly*", corresponding to fixed ranges of the continuous target force shared across tasks. For example, prompts include "Wipe the table *gently*.", "Push the box into the shelf *firmly*.", and "Pick up the box *strongly*."

B. Force-Aware RL-Based Whole-Body Control

The motion and contact-force references predicted by the VLA are executed by task-specific FCT controllers. We formulate each controller as an RL policy that outputs joint-position commands and adopt an asymmetric actor-critic architecture [40] for sim-to-real transfer.

Policy and Observation. At each timestep t, the actor and critic are conditioned on motion and contact-force references:

$$\begin{aligned} a_t &\sim \pi_\theta\Big(a_t \mid o_t^{\text{actor}},\, \tilde{\mathbf{p}}_t^{\text{ee}},\, \tilde{\mathbf{F}}_t^{\text{ref}}\Big), \\ V_\phi &= V_\phi\big(o_t^{\text{critic}},\, \mathbf{p}_t^{\text{ee}},\, \mathbf{F}_t^{\text{ref}}\big), \end{aligned} \tag{2}$$

where $\mathbf{p}_t^{\text{ee}}$ denotes the four end-effector motion targets and $\mathbf{F}_t^{\text{ref}}$ denotes the target contact force. The tilde $(\tilde{\cdot})$ denotes noisy reference information used to match deployment conditions.

The actor observation o_t^{actor} consists of deployable proprioceptive signals, including joint states, base velocities, action history, and measured joint torques. The critic observation o_t^{critic} includes additional privileged simulation states and TO references. Detailed observations are provided in Table I.

Action Space. The policy outputs joint-position offsets from a default configuration, $\mathbf{q}_t^{\text{cmd}} = \mathbf{q}^{\text{dflt}} + a_t$, which are converted to joint torques through a PD controller: $\mathbf{u}_t = \mathbf{K}_p(\mathbf{q}_t^{\text{cmd}} - \mathbf{q}_t) - \mathbf{K}_d\dot{\mathbf{q}}_t$. Here, $\mathbf{q}_t$ denotes the actuated joint positions. The policy learns to adjust these setpoints from motion and contact-force references, proprioception, and measured joint torques, implementing force regulation through the inner PD loop.

Force-Aware Training. The controller is trained with:

$$r_t = r_t^{\text{motion}} + \lambda_f\, r_t^{\text{force}} + \lambda_\tau\, r_t^{\text{torque}} + r_t^{\text{reg}}, \tag{3}$$

where r_t^{motion} tracks geometric references, including joint positions, base position and orientation, base linear ad angular velocities, and task-specific end-effector targets. r_t^{force} tracks the target contact force, r_t^{torque} tracks TO-derived reference joint torques, and r_t^{reg} regularizes the motion. Importantly, contact-force references are provided to the actor during both training and deployment, whereas reference torque provides physically grounded guidance only during training. We follow the domain randomization scheme of [37], including randomized dynamics, observation noise, control latency, and external perturbations. In addition, we apply Gaussian noise to the motion and force reference observations to improve robustness to prediction errors from the VLA policy.

TABLE I
OBSERVATION AND STATE SPACES OF THE FORCE-CONDITIONED CONTROLLER WITH TORQUE SUPERVISION (FCT).

Input	Symbol	Dim	Actor	Critic
Base Lin. Vel.	$\dot{\mathbf{p}}_t^{\text{b}}$	3	✓	✓
Base Ang. Vel.	$\boldsymbol{\omega}_t^{\text{b}}$	3	✓	✓
Projected Gravity	$\mathbf{g}_t$	3	✓	✓
History Motor Joint Pos.	$\mathbf{q}_{\text{hist}}^{\text{j}}$	200	✓	✓
Motor Joint Vel.	$\dot{\mathbf{q}}_t^{\text{j}}$	20	✓	✓
Ref. End-effector Pos.	$\hat{\mathbf{p}}_t^{\text{ee}}$	12	✓	✓
History Action	$\mathbf{a}_{\text{hist}}$	200	✓	✓
Upper Body Joint Torque	$\mathbf{u}_t$	8	✓	✓
Ref. Contact Force	$\hat{\mathbf{F}}_t^{\text{c}}$	2	✓	✓
Base Translation	$\mathbf{p}^{\text{b}}$	3		✓
Base Orientation	$\boldsymbol{\theta}_t^b$	3		✓
End-effector Pos.	$\mathbf{p}^{\text{e}}$	12		✓
Ref. Base Translation	$\hat{\mathbf{p}}^{\text{b}}$	3		✓
Ref. Base Orientation	$\hat{\boldsymbol{\theta}}_t^b$	3		✓
Ref. Base Lin. Vel.	$\hat{\dot{\mathbf{p}}}^{\text{b}}$	3		✓
Ref. Base Ang. Vel.	$\hat{\boldsymbol{\omega}}^{\text{b}}$	3		✓
Ref. Actuator Joint Pos.	$\hat{\mathbf{q}}^{\text{j}}$	20		✓
Ref. Upper Body Joint Torque	$\hat{\mathbf{u}}_t$	8		✓
Contact Force	$\mathbf{F}_t^{\text{c}}$	2		✓
PD Gains	$\mathbf{K}_p, \mathbf{K}_d$	40		✓
Motor Strength Scale	$\boldsymbol{\alpha}$	20		✓

Note that all the force related information including "Upper Body Joint Torque", "Ref. Upper Body Joint Torque", "Contact Force", "Ref. Contact Force" are not included in the motion-only controller (MO), and "Ref. Upper Body Joint Torque" is not included in the force-conditioned controller (FC).

Deployment. The policy runs at 200 Hz, with an internal PD loop at 1 kHz in simulation and 2 kHz on hardware. At deployment, no TO reference torques are required; the controller tracks the contact-force references predicted by the vision-language planner using onboard observations.

C. Whole-Body Trajectory Optimization for Force Supervision

Following the TO pipeline in Opt2Skill [37], we offline whole-body TO to generate supervision for FCT training. Given the robot model, task-specific end-effector motion targets, contact schedule, and prescribed contact wrenches, TO optimizes dynamically feasible whole-body state and joint-torque trajectories subject to floating-base dynamics, contact constraints friction, and joint and actuator limits. We solve the TO using a modified implementation of Crocoddyl and Pinocchio [41], [42], adapted to our contact modeling and constraint requirements. The optimized motion trajectory provides geometric references, while the optimized joint torques provide privileged supervision for FCT training; the prescribed interaction wrench defines the target force command. By varying the task configuration and prescribed force, we generate references across different interaction conditions. The resulting FCT controllers are subsequently rolled out across different task configurations and force conditions to collect multimodal training data for the VLA policy.

IV. RESULTS

A. Experimental Setup and Tasks

We evaluate the proposed force-aware VLA framework in simulation and real-world settings on a full-size humanoid robot Digit developed by Agility Robotics. The robot weighs

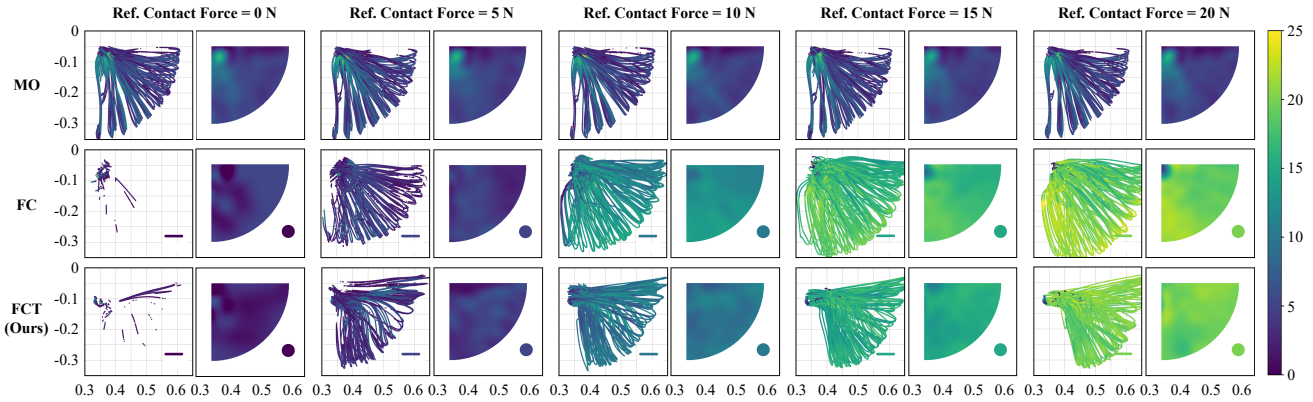

Fig. 3. **Contact force distributions for the surface wiping task under different target forces.** Each subplot shows scattered measured contact forces (left) across trials at global end-effector (x, y) positions, along with a Gaussian fit (right) summarizing the resulting distribution. Results are aggregated over 10×3 trials with varying motion directions and table heights. Axes indicate position in meters, and color encodes normal contact force in newtons. FC and FCT produce force levels that vary with the target, whereas MO produces similar forces across target conditions. FCT achieves lower average force-tracking error than FC, as shown in Table II.

approximately 48 kg and has 30 degrees of freedom (DoF), including 20 actuated joints.

We consider three daily manipulation tasks involving sustained physical interaction under varying contact conditions and object properties:

1) **Surface wiping.** The robot wipes a planar surface using repetitive back-and forth motions while maintaining contact. Appropriate interaction force varies with surface properties and task context.
2) **Shelf-box pushing.** The robot pushes a box toward a designated shelf wall and then regulates the contact force against it. Visual observations become partially occluded while physical interaction remains observable.
3) **Box pickup with varying weight.** The robot grasps and lifts visually identical boxes with different masses, requiring appropriate grasp forces for reliable and safe manipulation.

Together, these tasks provide different interaction conditions for evaluating force-aware whole-body control and vision-language planning. Example prompts are provided in Sec. III-A.

B. Force-Aware Whole-Body Control

We first evaluate Opt2VLA's low-level force regulation to examine how force-conditioned training affects contact-force regulation and whether TO-derived joint-torque supervision provides additional improvements.

Controller variants. We compare three variants trained with the same RL framework:

1) **Motion-only control (MO).** The controller tracks geometric motion references without target-force inputs or force-tracking objectives.
2) **Force-conditioned control (FC).** The controller additionally receives a target interaction force and is trained with a force-tracking objective, without reference joint-torque supervision.
3) **Force-conditioned control with torque supervision (FCT, Ours).** In addition to target-force conditioning, FCT uses TO-derived reference joint torques as privileged supervision during training to provide physically grounded guidance for force regulation.

We first evaluate the *surface wiping* task, where similar geometric motions are executed under different target interaction forces. Fig. 3 visualizes the measured contact-force distributions across target force levels and motion directions. Without target-force inputs, MO exhibits large force variability and produces similar distributions across the evaluated force levels. FC separates the contact force regimes but exhibits larger tracking errors, particularly at lower force levels. In contrast, FCT improves contact-force tracking accuracy and achieves lower average variance, as shown in Table II. These results show that explicit force conditioning enables different interaction regimes, while TO-derived torque supervision further improves force regulation accuracy and stability. The torque references are computed jointly with the reference motion under prescribed contact wrenches, subject to whole-body dynamics, contact constraints, and actuator limits. They therefore provide physically consistent joint-level guidance for realizing the desired interaction force.

The *shelf-box pushing* task exhibits similar but more pronounced differences. The end-effector pushes the box against the designated shelf wall and maintains the specified interaction force. Fig. 4 shows the measured force distributions across different target force levels. MO again produces similar interaction forces across conditions due to the absence of target-force commands, whereas FC separates the target regimes and produces distinct distributions, but exhibits poor force tracking accuracy. FCT achieves the most concentrated distributions and lowest force error (4.2 ± 5.5 N on average), as shown in Table II.

Finally, we evaluate the *box pickup* task across varying object weights. All variants receive identical geometric pickup trajectories, while FC and FCT additionally recieve target grasp-force commands spanning 3–20 N. As shown in Fig. 5, MO exhibits similar force behavior across object weights (Fig. 5(a)) and fails to explicitly regulate the resulting grasp force. In contrast, FC and FCT maintain a clear monotonic relationship between commanded and realized force across different weights (Figs. 5(b)–(i)). However, FC shows larger

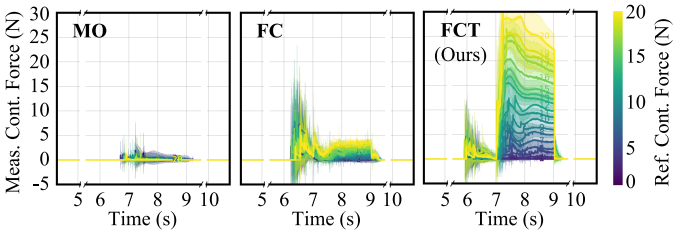

Fig. 4. **Measured push forces for the shelf-box pushing task under different target-force commands**. MO fails to separate force regimes, FC distinguishes force levels but exhibits large tracking error, while FCT achieves more concentrated force distributions and the lowest tracking error. Results at each target force level are aggregated across 10 target positions to evaluate force tracking over diverse motions, covering 20 target force levels and 200 trials in total.

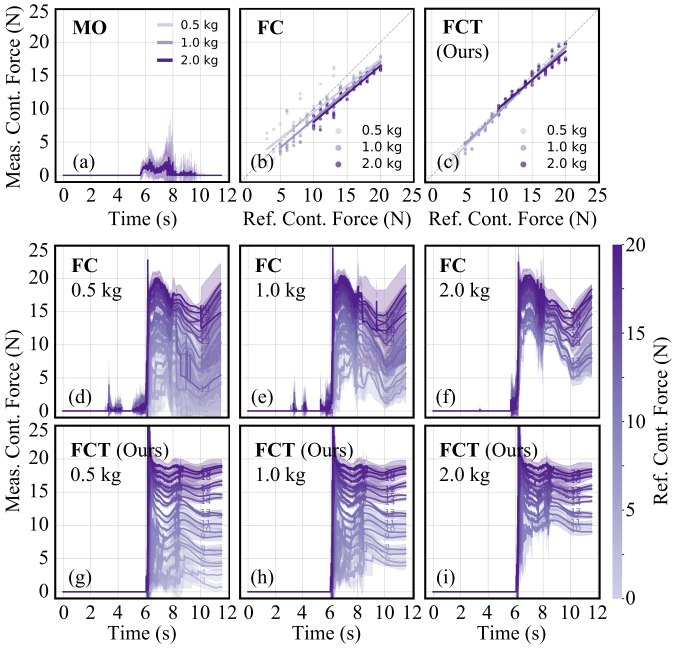

Fig. 5. **Force realization in the box pickup task.** (a) Measured grasp force trajectories over time for MO under different object weights. (b–c) Realized versus commanded grasp force for FC and FCT across object weights; points denote average lift-and-hold forces with linear fits per weight. (d–f) FC grasp force trajectories over time for different object weights and commanded force levels. (g–i) Corresponding force trajectories for FCT. The reference-force ranges are 3–20 N, 5–20 N, and 10–20 N for the 0.5 kg, 1.0 kg, and 2.0 kg boxes, respectively, with 10 trials per weight–force condition.

tracking errors, particularly for heavier objects. FCT achieves the most consistent force realization, achieving an average grasp-force error of 0.9 ± 1.0 N (Table II).

Across these tasks, force-conditioned policies produce contact forces that are consistent with the commanded force values, while TO-derived torque supervision further improves force tracking accuracy and consistency across the evaluated contact conditions.

C. Vision-Language Planning with Explicit Force Commands

We evaluate closed-loop Opt2VLA execution using nine multi-task VLA configurations that predict motion and contact-force references and share the same task-specific FCT controllers. The configurations differ in torque encoding, observation history, auxiliary future-torque supervision, and backbone fine-tuning. Each variant is evaluated on ten reference-defined scenes per task under *gentle*, *firm*, and *strong* prompts, yielding 90 episodes. Scene appearances use observed training-appearance families. Scenes, prompts, and inference settings are shared across variants.

TABLE II
FORCE TRACKING ERRORS UNDER DIFFERENT CONTACT FORCE REFERENCES (MAE $\pm$ STD IN N; LOWER IS BETTER).

Task	Ref. Force	5 N	10 N	15 N	20 N	Avg.
Surface Wiping	FC	3.1 ± 2.4	2.6 ± 2.5	2.5 ± 1.9	1.5 ± 1.8	2.4 ± 2.9
	FCT	$\mathbf{2.6 \pm 2.6}$	$\mathbf{1.7 \pm 2.1}$	$\mathbf{1.3 \pm 1.5}$	$\mathbf{1.0 \pm 1.2}$	$\mathbf{1.7 \pm 2.1}$
Shelf Pushing	FC	4.7 ± 0.6	9.2 ± 1.1	13.7 ± 1.1	17.1 ± 1.4	11.2 ± 4.8
	FCT	$\mathbf{3.9 \pm 1.2}$	$\mathbf{3.5 \pm 3.4}$	$\mathbf{2.5 \pm 3.5}$	$\mathbf{7.0 \pm 6.7}$	$\mathbf{4.2 \pm 5.5}$
Box Pickup	FC	1.0 ± 1.2	1.6 ± 0.9	2.5 ± 0.6	3.5 ± 0.6	2.3 ± 1.4
	FCT	$\mathbf{0.7 \pm 0.5}$	$\mathbf{0.5 \pm 0.5}$	$\mathbf{0.6 \pm 0.7}$	$\mathbf{1.6 \pm 1.1}$	$\mathbf{0.9 \pm 1.0}$

TABLE III
CLOSED-LOOP OPT2VLA ABLATION: JOINT SUCCESS RATES (%). TASK COLUMNS AVERAGE THE THREE FORCE PROMPTS.

Variant	H_s	Torque	Aux.	Train	Box pickup	Shelf pushing.	Surface wiping	Overall
Baseline (C0)	1	–	–	T-H	33.3	53.3	46.7	44.4
+ future-τ (C1)	1	–	✓	T-H	33.3	86.7	63.3	61.1
State torque (C2)	1	State, 1	–	T-H	50.0	46.7	50.0	48.9
+ future-τ (C3)	1	State, 1	✓	T-H	60.0	73.3	66.7	66.7
State history (C4)	10	State, 10	–	T-H	6.7	6.7	33.3	15.6
+ future-τ (C5)	10	State, 10	✓	T-H	13.3	66.7	33.3	37.8
Token history (C6)	10	Token, 10	–	T-H	53.3	13.3	100.0	55.6
+ future-τ (C7, Ours)	10	Token, 10	✓	T-H	76.7	70.0	100.0	82.2
Full FT (C9)	10	Token, 10	✓	Full	76.7	23.3	100.0	66.7

$n = 30$ per task and 90 overall. Underlines mark the best rates, including ties.
C7 is the default Opt2VLA planner configuration; all variants use the same task-specific FCT controllers.
H_s: state-history length. Torque entries give the encoder and history length: State, concatenation with state; Token, a dedicated torque encoder. Ten-sample histories use 200-ms spacing. Aux.: future measured-torque prediction.
T-H: frozen VLM with connectors, action model, and VLM normalization trained; Full also trains the VLM.
Each row represents one VLA model evaluated across all three tasks.

Evaluation Criteria. We assess task completion, prompt-band agreement, and mean-force agreement. Pickup requires the lowest point of the box to clear the tabletop by at least 5 cm; shelf-box pushing requires reaching the designated right-side wall; and wiping requires a continuous tabletop-contact stroke of at least 5 cm within 30° of the vertical wiping axis. Prompt-band agreement requires the mean commanded force during the active-command window to fall within the training-defined force range associated with the instruction, allowing a 2 N tolerance at each boundary. The resulting acceptance bands are $[0, 9]$ N, $(5, 16]$ N, and $(12, 22]$ N for *gentle*, *firm*, and *strong*, respectively. Mean-force agreement requires the mean commanded and measured forces to differ by at most 5 N during a phase in which both are active. Joint success requires all three criteria to be satisfied within the same attempt. Joint success rate (SR) is computed over all valid episodes.

Language-Conditioned Motion–Force Execution. Table III summarizes joint success. Completing the geometric task does not necessarily satisfy the requested force level: the kinematic-input variant (C0) achieves a task-only SR of 85.6%, which decreases to 45.6% when prompt-band agreement is also required. Opt2VLA (C7) achieves 88.9% under this

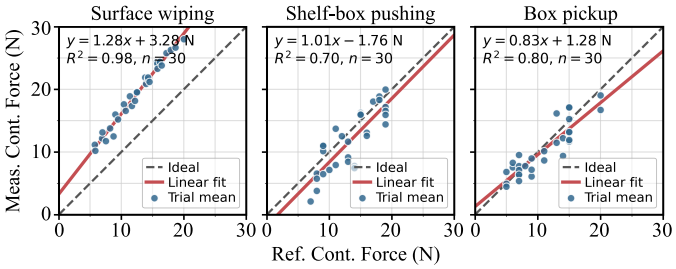

Fig. 6. **Hardware force tracking of Opt2VLA's FCT controller.** (a) surface wiping, (b) shelf-box pushing, and (c) box pickup. Markers show per-trial mean, baseline-corrected force during steady contact; dashed lines indicate ideal tracking and solid lines show least-squares fits.

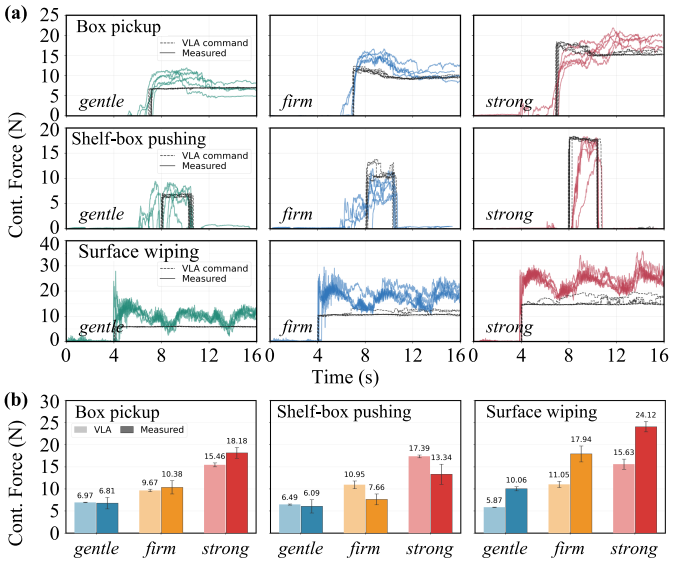

Fig. 7. **End-to-end Opt2VLA force modulation on hardware.** (a) VLA commands (dashed) and measured forces (solid). (b) Mean commanded and measured forces over the task-specific evaluation phase. Results include five trials per task and prompt.

combined criterion and the highest overall joint SR of 82.2%, demonstrating stronger agreement between task execution, language-specified force levels, and realized interaction.

Torque-Aware Planning. The benefit of interaction feedback depends on its representation and supervision. With the same observation histories and no auxiliary loss, a dedicated torque encoder (C6) achieves a joint SR of 55.6%, compared with 15.6% for native state concatenation (C4). Adding future-torque supervision (C7) further improves the joint SR, showing that learning to predict future joint torques helps the VLA generate more effective motion and force commands.

The improvement is task-dependent: C1 achieves the highest *shelf-box pushing* SR of 86.7%. Full-backbone fine-tuning (C9) matches C7 on *box pickup* and *surface wiping*, but performs worse on *shelf-box pushing*, reducing overall joint SR to 66.7%. Overall, Opt2VLA (C7) achieves the highest joint SR across the evaluated tasks using a dedicated torque encoder, observation histories, and auxiliary future-torque supervision with a frozen vision-language backbone.

D. Hardware Evaluation

RL force tracking. We evaluate force tracking by Opt2VLA's FCT controller over 30 hardware trials per task, as shown in Fig. 6. For surface wiping, the normal-force error is computed over the detected sustained-contact phase. With a spanning reference forces of 5.0–20.0 N, the controller achieves a mean absolute error (MAE) of 7.0 N (95% bootstrap CI: 6.5–7.4 N). The measured force remained correlated with the reference ($R^2 = 0.98$), showing consistent modulation across commanded force levels.

For shelf-box pushing, force is evaluated over the sustained-contact phase. the controller achieves an MAE of 3.9 N (95% bootstrap CI: 3.6–4.3 N). The realized force varied approximately linearly with the reference with $R^2 = 0.70$, indicating that the controller reproduced the overall trend of the commanded force across force levels. For box pickup, we evaluate the steady-state force tracking over the final 2 s of the holding phase. The controller achieves an MAE of 1.7 N (95% bootstrap CI: 1.4–2.1 N). The measured force exhibits a strong relationship with the reference force, with a calibration slope of 0.83 and $R^2 = 0.80$.

Overall, FCT preserves the commanded force ordering and consistent force modulation after sim-to-real transfer, although the absolute tracking accuracy varies across tasks.

End-to-End Opt2VLA execution. We evaluate the complete Opt2VLA system over 45 hardware trials, using the same VLA checkpoint across all three tasks, with five trials per task under *gentle*, *firm*, and *strong* prompts (Fig. 7). The mean VLA-issued force commands fall within the intended force regimes across all three tasks. Surface wiping exhibits a larger command-to-measurement gap, consistent with the tracking bias observed in the RL-only hardware evaluation. This suggests that low-level force-tracking error contributes to the gap even when the VLA selects the intended force level. Overall, these results demonstrate language-conditioned force modulation during closed-loop hardware execution.

V. Conclusion

We present Opt2VLA, a force-aware VLA framework that introduces explicit force commands as the interface between vision-language planning and humanoid whole-body control. The policy jointly predicts geometric motion goals and contact-force references, which are tracked by an RL-based controller trained to regulate motion and interaction force. Whole-body TO which prescribed contact wrenches provides dynamically feasible motion and joint-torque references for physically grounded controller training and scalable multimodal data generation. Experiments on three humanoid whole-body manipulation tasks show that explicit force conditioning enables distinct interaction regimes beyond geometric tracking alone, while TO-derived torque supervision further improves force-tracking accuracy and consistency. Closed-loop VLA evaluation further shows that by adding torque history via a separate encoder and adding an auxiliary future torque loss, we can achieve much better force command generation and force command tracking. Hardware experiments demonstrate the sim-to-real transfer of the Opt2VLA framework to a fill-size humanoid robot.

REFERENCES

[1] J. A. Barreiros, A. Ö. Önol, M. Zhang, S. Creasey, A. Goncalves, A. Beaulieu, A. Bhat, K. M. Tsui, and A. Alspach, "Learning contact-rich whole-body manipulation with example-guided reinforcement learning," *Science Robotics*, vol. 10, no. 105, p. eads6790, 2025.

[2] Z. Gu, J. Li, W. Shen, W. Yu, Z. Xie, S. McCrory, X. Cheng, A. Shamsah, R. Griffin, C. K. Liu, *et al.*, "Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning," *IEEE/ASME Transactions on Mechatronics*, vol. 31, no. 2, pp. 2300–2330, 2026.

[3] C. Khazoom, S. Hong, M. Chignoli, E. Stanger-Jones, and S. Kim, "Tailoring solution accuracy for fast whole-body model predictive control of legged robots," *IEEE Robotics and Automation Letters*, 2024.

[4] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "Humanplus: Humanoid shadowing and imitation from humans," in *Conference on Robot Learning (CoRL)*, 2024.

[5] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, *et al.*, "Openvla: An open-source vision-language-action model," *arXiv preprint arXiv:2406.09246*, 2024.

[6] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid, *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," in *Conference on Robot Learning*. PMLR, 2023, pp. 2165–2183.

[7] J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding, L. Fan, Y. Fang, D. Fox, F. Hu, S. Huang, *et al.*, "Gr00t n1: An open foundation model for generalist humanoid robots," *arXiv preprint arXiv:2503.14734*, 2025.

[8] P. Intelligence, A. Amin, R. Aniceto, A. Balakrishna, K. Black, K. Conley, G. Connors, J. Darpinian, K. Dhabalia, J. DiCarlo, *et al.*, "$\pi^{\star}_{0.6}$: a vla that learns from experience," *arXiv preprint arXiv:2511.14759*, 2025.

[9] P. Ding, J. Ma, X. Tong, B. Zou, X. Luo, Y. Fan, T. Wang, H. Lu, P. Mo, J. Liu, *et al.*, "Humanoid-vla: Towards universal humanoid control with visual integration," *arXiv preprint arXiv:2502.14795*, 2025.

[10] H. Yuan, Y. Bai, Y. Fu, B. Zhou, Y. Feng, X. Xu, Y. Zhan, B. F. Karlsson, and Z. Lu, "Being-0: A humanoid robotic agent with vision-language models and modular skills," *arXiv preprint arXiv:2503.12533*, 2025.

[11] H. Jiang, J. Chen, Q. Bu, L. Chen, M. Shi, Y. Zhang, D. Li, C. Suo, C. Wang, Z. Peng, *et al.*, "Wholebodyvla: Towards unified latent vla for whole-body loco-manipulation control," *arXiv preprint arXiv:2512.11047*, 2025.

[12] Z. Luo, Y. Yuan, T. Wang, C. Li, F. Castañeda, S. Chen, Z.-A. Cao, J. Li, D. Minor, Q. Ben, *et al.*, "Sonic: Supersizing motion tracking for natural humanoid whole-body control," *Science Robotics*, vol. 11, no. 117, p. eaed4592, 2026.

[13] J. Yu, H. Liu, Q. Yu, J. Ren, C. Hao, H. Ding, G. Huang, G. Huang, Y. Song, P. Cai, *et al.*, "Forcevla: Enhancing vla models with a force-aware moe for contact-rich manipulation," *arXiv preprint arXiv:2505.22159*, 2025.

[14] Z. Zhang, H. Xu, Z. Yang, C. Yue, Z. Lin, H.-a. Gao, Z. Wang, and H. Zhao, "Ta-vla: Elucidating the design space of torque-aware vision-language-action models," *arXiv preprint arXiv:2509.07962*, 2025.

[15] Y. Zhang, S. Ma, L. Yang, Y. Li, C. Hao, H. Chi, D. We, Q. Yu, and D. Hou, "Fwbc-vla: Force-aware whole-body compensation for contact-rich loco-manipulation," *arXiv preprint arXiv:2609.03889*, 2026.

[16] Q. Rouxel, S. Ivaldi, and J.-B. Mouret, "Multi-contact whole-body force control for position-controlled robots," *IEEE Robotics and Automation Letters*, vol. 9, no. 6, pp. 5639–5646, 2024.

[17] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa, "Biped walking pattern generation by using preview control of zero-moment point," in *2003 IEEE international conference on robotics and automation (Cat. No. 03CH37422)*, vol. 2. IEEE, 2003, pp. 1620–1626.

[18] P. M. Wensing, M. Posa, Y. Hu, A. Escande, N. Mansard, and A. Del Prete, "Optimization-based control for dynamic legged robots," *IEEE Transactions on Robotics*, vol. 40, pp. 43–63, 2023.

[19] J. Koenemann, A. Del Prete, Y. Tassa, E. Todorov, O. Stasse, M. Bennewitz, and N. Mansard, "Whole-body model-predictive control applied to the hrp-2 humanoid," in *2015 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2015, pp. 3346–3351.

[20] A. Adu-Bredu, G. Gibson, and J. Grizzle, "Exploring kinodynamic fabrics for reactive whole-body control of underactuated humanoid robots," in *2023 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2023, pp. 10 397–10 404.

[21] X. B. Peng, P. Abbeel, S. Levine, and M. Van de Panne, "Deepmimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Transactions On Graphics (TOG)*, vol. 37, no. 4, pp. 1–14, 2018.

[22] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, and K. Sreenath, "Real-world humanoid locomotion with reinforcement learning," *Science Robotics*, vol. 9, no. 89, p. eadi9579, 2024.

[23] J. Dao, H. Duan, and A. Fern, "Sim-to-real learning for humanoid box loco-manipulation," in *2024 IEEE Int. Conf. Robot. Autom. (ICRA)*. IEEE, 2024, pp. 16 930–16 936.

[24] X. Cheng, Y. Ji, J. Chen, R. Yang, G. Yang, and X. Wang, "Expressive whole-body control for humanoid robots," *arXiv preprint arXiv:2402.16796*, 2024.

[25] P. Dugar, A. Shrestha, F. Yu, B. van Marum, and A. Fern, "Learning multi-modal whole-body control for real-world humanoid robots," in *Proceedings of the AAAI Symposium Series*, vol. 7, no. 1, 2025, pp. 650–657.

[26] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi, "Learning human-to-humanoid real-time whole-body teleoperation," in *2024 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2024, pp. 8944–8951.

[27] G. B. Margolis, M. Wang, N. Fey, and P. Agrawal, "Softmimic: Learning compliant whole-body control from examples," *arXiv preprint arXiv:2510.17792*, 2025.

[28] F. Jenelten, J. He, F. Farshidian, and M. Hutter, "Dtc: Deep tracking control," *Science Robotics*, vol. 9, no. 86, p. eadh5401, 2024.

[29] Y. Fuchioka, Z. Xie, and M. Van de Panne, "Opt-mimic: Imitation of optimized trajectories for dynamic quadruped behaviors," in *International Conference on Robotics and Automation*, 2023, pp. 5092–5098.

[30] M. Dai, W. D. Compton, J. Li, L. Yang, and A. D. Ames, "Walk the planc: Physics-guided rl for agile humanoid locomotion on constrained footholds," *arXiv preprint arXiv:2601.06286*, 2026.

[31] J. Li, L. Wu, S. A. Esteban, L. Yang, J. Drgona, and A. D. Ames, "Accelerating and scaling mpc-guided reinforcement learning for humanoid locomotion and manipulation," *arXiv preprint arXiv:2606.05687*, 2026.

[32] R. Batke, F. Yu, J. Dao, J. Hurst, R. L. Hatton, A. Fern, and K. Green, "Optimizing bipedal maneuvers of single rigid-body models for reinforcement learning," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. IEEE, 2022, pp. 714–721.

[33] D. Marew, N. Perera, S. Yu, S. Roelker, and D. Kim, "A biomechanics-inspired approach to soccer kicking for humanoid robots," in *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*. IEEE, 2024, pp. 722–729.

[34] E. Chaikovskaya, I. Minashina, V. Litvinenko, E. Davydenko, D. Makarov, Y. Danik, and R. Gorbachev, "Benchmarking the full-order model optimization based imitation in the humanoid robot reinforcement learning walk," in *2023 21st International Conference on Advanced Robotics (ICAR)*. IEEE, 2023, pp. 206–211.

[35] L. Krishna, G. A. Castillo, U. A. Mishra, A. Hereid, and S. Kolathaya, "Linear policies are sufficient to realize robust bipedal walking on challenging terrains," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2047–2054, 2022.

[36] Z. Olkin, K. Li, W. D. Compton, and A. D. Ames, "Chasing stability: Humanoid running via control lyapunov function guided reinforcement learning," *arXiv preprint arXiv:2509.19573*, 2025.

[37] F. Liu, Z. Gu, Y. Cai, Z. Zhou, H. Jung, J. Jang, S. Zhao, S. Ha, Y. Chen, D. Xu, *et al.*, "Opt2skill: Imitating dynamically-feasible whole-body trajectories for versatile humanoid loco-manipulation," *IEEE Robotics and Automation Letters*, 2025.

[38] X. Xu, Y. Zhang, Y.-L. Li, L. Han, and C. Lu, "Humanvla: Towards vision-language directed object rearrangement by physical humanoid," *Advances in Neural Information Processing Systems*, vol. 37, pp. 18 633–18 659, 2024.

[39] H. Xue, X. Huang, D. Niu, Q. Liao, T. Kragerud, J. T. Gravdahl, X. B. Peng, G. Shi, T. Darrell, K. Sreenath, *et al.*, "Leverb: Humanoid whole-body control with latent vision-language instruction," *arXiv preprint arXiv:2506.13751*, 2025.

[40] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," *arXiv preprint arXiv:1710.06542*, 2017.

[41] C. Mastalli, R. Budhiraja, W. Merkt, G. Saurel, B. Hammoud, M. Naveau, J. Carpentier, L. Righetti, S. Vijayakumar, and N. Mansard, "Crocoddyl: An efficient and versatile framework for multi-contact optimal control," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2020, pp. 2536–2542.

[42] J. Carpentier, G. Saurel, G. Buondonno, J. Mirabel, F. Lamiraux, O. Stasse, and N. Mansard, "The pinocchio c++ library: A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives," in *2019 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2019, pp. 614–619.

VI. APPENDIX

A. Reward Design and Domain Randomization

TABLE IV
REWARD COMPONENTS AND WEIGHTS.

Category	Term	Expression	Weight
Task Reward	Joint Pos.	$\exp(-5\|\hat{\mathbf{q}}_t^{\mathrm{j}} - \mathbf{q}_t^{\mathrm{j}}\|_2^2)$	3
	Base Pos.	$\exp(-20\|\hat{\mathbf{p}}_t^{\mathrm{b}} - \mathbf{p}_t^{\mathrm{b}}\|_2^2)$	3
	Base Ori.	$\exp(-50\|\hat{\boldsymbol{\theta}}_t^{\mathrm{b}} - \boldsymbol{\theta}_t^{\mathrm{b}}\|_2^2)$	3
	Base Lin. Vel.	$\exp(-2\|\hat{\dot{\mathbf{p}}}_t^{\mathrm{b}} - \dot{\mathbf{p}}_t^{\mathrm{b}}\|_2^2)$	3
	Base Ang. Vel.	$\exp(-0.5\|\hat{\boldsymbol{\omega}}_t^{\mathrm{b}} - \boldsymbol{\omega}_t^{\mathrm{b}}\|_2^2)$	3
	End-effector Pos.	$\exp(-20\|\hat{\mathbf{p}}_t^{\mathrm{ee}} - \mathbf{p}_t^{\mathrm{ee}}\|_2^2)$	3
	Joint Torque	$\exp(-0.01\|\hat{\mathbf{u}}_t - \mathbf{u}_t\|_2^2)$	2
	Contact Force	$\exp(-0.05\|\hat{\mathbf{F}}_t^{\mathrm{c}} - \mathbf{F}_t^{\mathrm{c}}\|_1)$	2
Penalty Cost	Action Rate	$\|\mathbf{a}_t - 2\mathbf{a}_{t-1} + \mathbf{a}_{t-2}\|_2^2$	-3
	Torques	$\|\mathbf{u}_t / \mathbf{u}_{\text{limit}}\|_2^2$	-0.3
	Joint Acc.	$\|\ddot{\mathbf{q}}_t^{\mathrm{j}}\|_2^2$	-10^{-5}

Note that the "Joint Torque" and "Contact Force" rewards are not included in the MO controller, and "Joint Torque" reward is not included in the FC controller.

TABLE V
DOMAIN RANDOMIZATION PARAMETERS.

Category	Parameter	Type	Range / Std
Obs.	Ref. End-effector Pos. (m)	Additive (Gauss)	$\sigma = 0.02$
	Ref. Contact Force (N)	Additive (Gauss)	$\sigma = 2.0$
	Joint Pos. (rad)	Additive (Gauss)	$\sigma = 0.0875$
	Joint Vel. (rad/s)	Additive (Gauss)	$\sigma = 0.075$
	Base Lin. Vel. (m/s)	Additive (Gauss)	$\sigma = 0.075$
	Base Ang. Vel. (rad/s)	Additive (Gauss)	$\sigma = 0.075$
	Gravity Proj. (m/s^2)	Additive (Gauss)	$\sigma = 0.0375$
Delays	Action Delay (s)	Uniform	$[0.0, 0.02]$
Motor	Motor Strength	Scaling (Uniform)	$[0.95, 1.05]$
	Kp/Kd Factor	Scaling (Uniform)	$[0.9, 1.1]$
Reset	Joint pos. init. (rad)	Additive (Gauss)	$\sigma = 0.03$
	Joint vel. init. (rad/s)	Additive (Gauss)	$\sigma = 0.06$
	Root vel. init. (m/s)	Additive (Gauss)	$\sigma = 0.05$
Env.	Gravity (m/s^2)	Scaling (Uniform)	$[0.9, 1.1]$
	Wipe: Table height offset (m)	Additive (Uniform)	$[-0.01, 0.01]$

B. Whole-Body Trajectory Optimization for Force Supervision

We use whole-body TO offline to generate physically grounded supervision for force-aware planning and control. Given task-specific motion targets and prescribed contact wrenches, TO generates dynamically feasible trajectories that satisfy whole-body dynamics and contact constraints. This provides geometric motion and joint-torque references that are physically consistent with the desired interaction forces, which are difficult to obtain from kinematic planning or human demonstrations.

We consider a standard floating-base humanoid model with generalized coordinates $\mathbf{q}$ and velocities $\mathbf{v}$, governed by the equations of motion:

$$\mathbf{M}(\mathbf{q})\dot{\mathbf{v}} + \mathbf{C}(\mathbf{q}, \mathbf{v}) = \mathbf{B}\mathbf{u} + \mathbf{J}_{\mathrm{c}}^{\top}\mathbf{F}_{\mathrm{c}}, \tag{4}$$

where $\mathbf{M}(\mathbf{q})$ is the joint-space mass matrix, $\mathbf{C}(\mathbf{q}, \mathbf{v})$ collects Coriolis, centrifugal, and gravitational terms, $\mathbf{B}$ is the actuation matrix, $\mathbf{u}$ denotes joint torques, $\mathbf{J}_{\mathrm{c}}(\mathbf{q})$ is the contact Jacobian, and $\mathbf{F}_{\mathrm{c}}$ denotes contact wrench forces.

Over a finite horizon, TO optimizes whole-body motion and joint torques under prescribed interaction wrenches:

$$\min_{\mathbf{x},\mathbf{u}} \quad \sum_{k=0}^{N-1} \Big(\|\mathbf{y}[k] - \hat{\mathbf{y}}[k]\|_Q^2 + \|\mathbf{u}[k]\|_R^2 \Big) + \|\mathbf{y}[N] - \hat{\mathbf{y}}[N]\|_{Q_f}^2 \tag{5a}$$

subject to

$$\text{(Dynamics)} \quad \begin{cases} \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{B}\mathbf{u} + \mathbf{J}_\mathrm{c}^\top \mathbf{F}_\mathrm{c} \\ \mathbf{M}(\mathbf{q})\dot{\mathbf{q}}^+ - \mathbf{M}(\mathbf{q})\dot{\mathbf{q}}^- = \mathbf{J}_\mathrm{c}^\top \mathbf{\Lambda} \end{cases} \tag{5b}$$

$$\text{(Contact)} \quad \begin{cases} \mathbf{J}_\mathrm{c}\ddot{\mathbf{q}} + \dot{\mathbf{J}}_\mathrm{c}\dot{\mathbf{q}} = 0 \\ \mathbf{J}_\mathrm{c}\dot{\mathbf{q}}^+ = 0 \end{cases} \tag{5c}$$

$$\text{(Limits)} \quad \mathbf{q}^\mathrm{j} \in \mathcal{J}, \mathbf{u} \in \mathcal{T} \tag{5d}$$

$$\text{(Friction)} \quad \mathbf{F}_\mathrm{c} \in \mathcal{F}(\mu, \mathbf{q}). \tag{5e}$$

Here $\mathbf{y} = \Phi(\mathbf{q})$ denotes task-space quantities such as end-effector pose. we solve the optimization using a differential dynamic programming (DDP)-based method implemented in Crocoddyl and Pinocchio [41], [42].

C. Hardware Force-Tracking Evaluation

Force-sensing setup and preprocessing. External force sensors were used only to evaluate hardware force tracking and were not provided to the control policy. For surface wiping, the vertical normal component F_z was obtained from a Bertec Force Plate (40 cm $\times 60$ cm). For shelf-box pushing and box pickup, force was measured using a SensX 160 force sensor (65 mm $\times 25$ mm). All signals were baseline-corrected. Filtering was used only to identify contact-phase boundaries; all force-tracking metrics were computed from the unfiltered, baseline-corrected measurements.

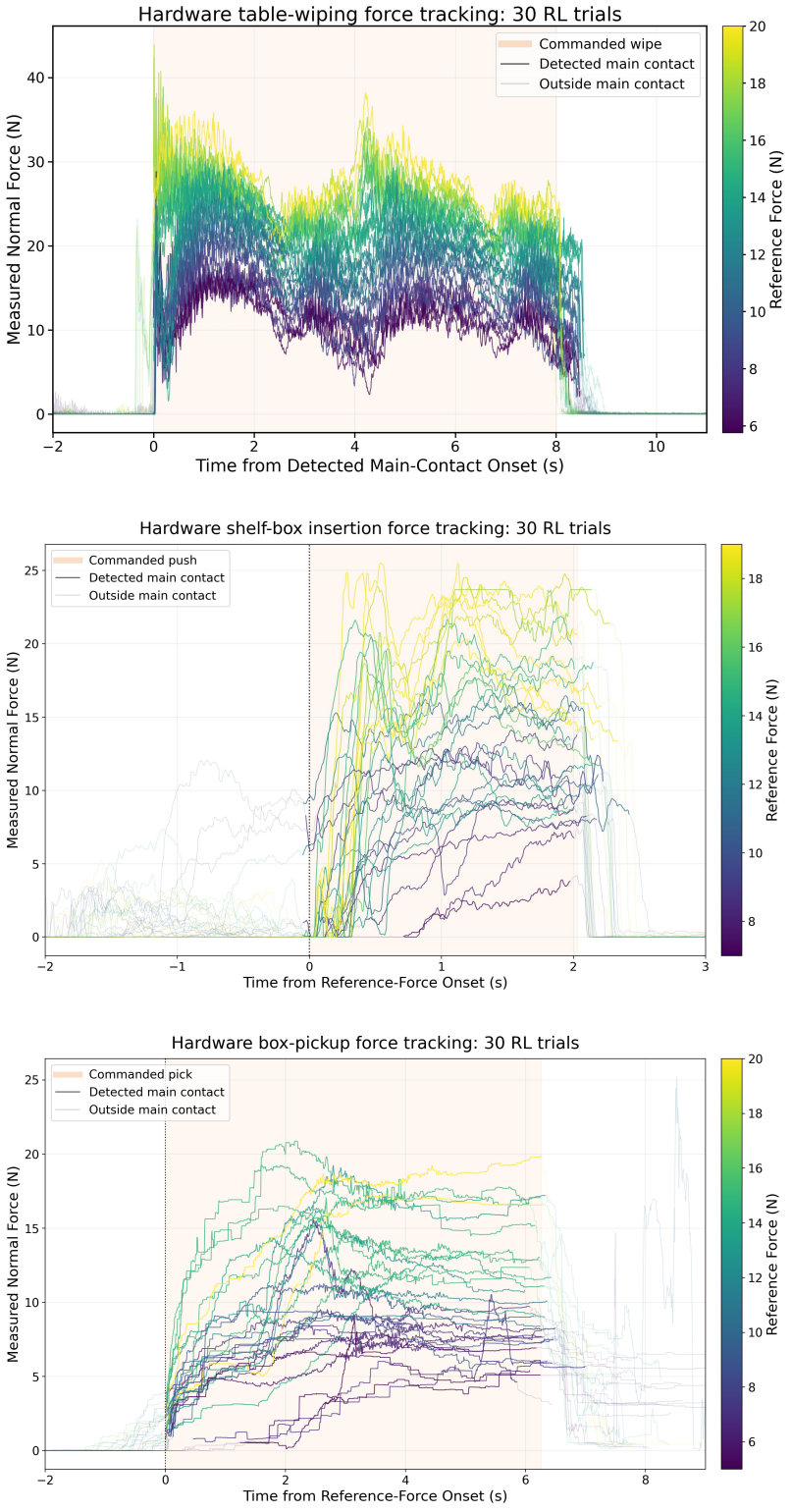

Fig. 8. **Hardware force trajectories across all trials.** Baseline-corrected normal-force trajectories are shown for 30 trials per task. Trajectory color indicates the reference force. Opaque segments denote the sustained-contact interval used for analysis, faded segments show measurements outside main contact, and orange shading indicates the commanded-force phase. Differences in trace smoothness reflect the sensing modalities used across tasks.

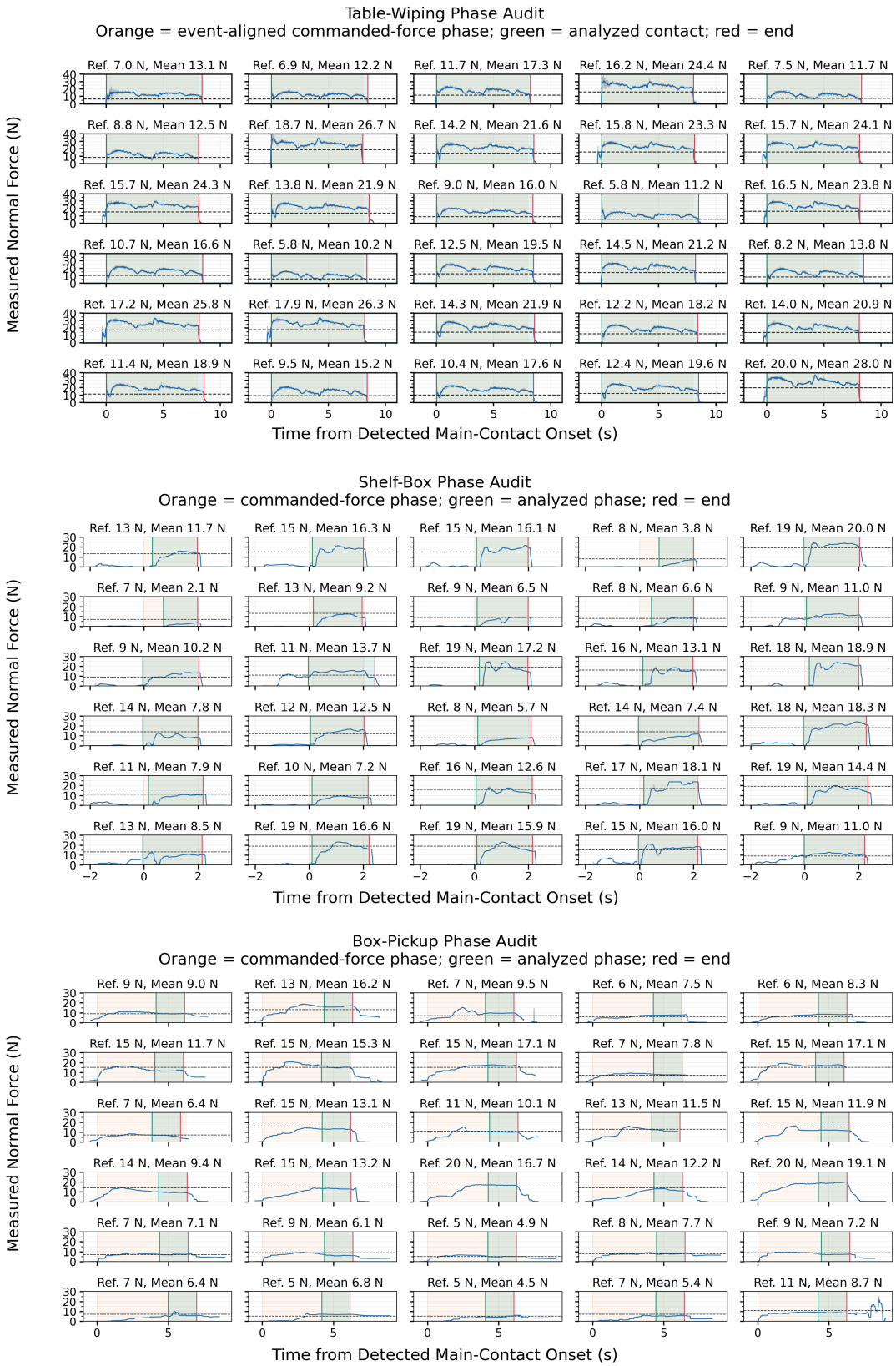

Fig. 9. **Contact-phase detection for the hardware trials.** Each subplot presents one trial. Gray curves show the unfiltered, baseline-corrected normal force; blue curves show the filtered signal used only for phase-boundary detection, and dashed black lines indicate the reference force. Orange and green shading shows the commanded-force phase and the retained analysis interval, respectively. The green and red vertical lines indicate the beginning and end of the analyzed interval.

D. Closed-Loop VLA Simulation Protocol

Evaluation design. We evaluate each variant using a single $40{,}000$-step checkpoint shared across all three tasks, with a fixed task-specific FCT controller. Each task uses ten reference-defined scenes and three force-level prompts: 30 episodes per task and 90 episodes per variant. Scene initialization, prompts, low-level controllers, and inference settings are shared across variants. Simulator and inference seeds are fixed to 123 and 0, respectively; the three prompts for a given scene use the same initial reset.
Scene and robot initialization. Reference trajectories specify support and object placement, nominal robot base pose, actuated-joint state, and initial velocities; the XML supplies remaining scene geometry and robot coordinates. Shelf scenes contain a fixed cabinet and a movable box.

Robot-reset perturbations add zero-mean Gaussian joint-position noise with a standard deviation of 0.03 rad. Initial base linear velocity, base angular velocity, and actuated-joint velocity are sampled from zero-mean Gaussian distributions with standard deviations 0.05 m/s, 0.05 rad/s, and 0.06 rad/s, respectively. Base position and orientation are not independently randomized; other domain randomization is disabled.
Language and appearance conditions. The prompt templates are *Pick up the box* {*gently/firmly/strongly*}., *Push the box into the shelf* {*gently/firmly/strongly*}., and *Wipe the table with* {*gentle/firm/strong*} *pressure.*, One descriptor is substituted for each episode without additional direction qualifiers.

Table VI lists the appearance configurations. Each appearance id assigned to two reference scenes and reused across all three force prompts and VLA variants. Floors and physical geometry remain unchanged.

TABLE VI
EVALUATION APPEARANCE CONFIGURATIONS. PICKUP AND SHELF ENTRIES SHOW SUPPORT/BOX PAIRS; WIPING ENTRIES DESCRIBE THE TABLE APPEARANCE. EACH APPEARANCE IS ASSIGNED TO TWO REFERENCE SCENES.

Task	Appearance configurations
Box pickup	White/yellow; yellow/white; blue/white; tan/blue; tan/yellow
Shelf pushing	White/yellow; yellow/blue; tan/blue; blue/tan; gray/yellow
Table wiping	White; gray-metal (textured); wood (textured); blue; yellow

Execution and observations. MuJoCo/MJX runs batches of four environments with native 320×240 chest-camera rendering. Physics integration, low-level control, issued VLA commands, ad VLA inference run at $1{,}000$, 200, 50, and 5 Hz, respectively. Each variant uses its trained observation schema; State and torque histories contain ten samples spaced 200 ms apart; unavailable initial history is filled by repeating the first observation. Future-torque targets are used only for auxiliary training and are not inference inputs.

Before VLA control begins, the task-specific FCT controller follows reference commands for a 1.8-s initialization period, which is excluded from scoring. Episode horizons are 20 s for pickup, 15 s for shelf-box pushing, and the reference endpoint (15.3–15.6 s) for wiping, subject to early termination.
Task-success criteria. Pickup succeeds when the lowest point of the oriented box collision geometry clears the tabletop by at least 5 cm, without a tilt constraint. Shelf-box pushing succeeds when the box reaches the designated right-wall, defined by wall overlap in world X/Z and a signed wall gap in $[-5, +0.01]$ mm. Wiping succeeds when the robot produces one continuous tabletop-contact stroke of at least 5 cm within 30° of the reset-heading forward–backward axis; either direction qualifies. Contact loss separates strokes, and a 1-cm reversal hysteresis separates successive wiping motions. No task requires goal-state dwell.
Force intervals and gates. Both hands are evaluated for pickup, while only the right hand is evaluated for shelf-box pushing and wiping. A command-active window starts when the maximum issued force over the evaluated hands exceeds 0.5 N continuously for 0.1 s, and ends when it remains at or below 0.5 N for 0.2 s. Within this window, a dual-active phase starts when all evaluated issued and measured force channels exceed 0.5 N continuously for 0.1 s. It ends when any channel remains at or below 0.5 N for 0.2 s, or when the command-active window ends. Detected boundaries are backdated to the beginning of the qualifying sequence; shorter interruptions remain included.

Prompt-band agreement requires each hand's mean issued force over the command-active window to lie within the force range associated with the language instruction. Starting from the nominal training ranges, we allow a 2 N tolerance at each boundary, resulting in acceptance bands of $[0, 9]$ N, $(5, 16]$ N, and $(12, 22]$ N for *gentle*, *firm*, and *strong*, respectively. Mean-force agreement passes when the mean issued and measured forces differ by at most 5 N per hand during a dual-active phase; both pickup hands must satisfy this criterion within the same phase. All means are computed at the controller-rate. Reference-trajectory forces are not used as scoring targets. Episodes without a valid command-active windows or dual-active phase cannot satisfy the corresponding force criterion.
Repeated attempts and joint success. Each command-active window defines an attempt. It's task outcome is evaluated until the next command window begins or active control ends. Force criteria use only the command and dual-active intervals associated with that attempt. An episode is counted as jointly successful if a single attempt satisfies task completion, prompt-band

agreement, and mean-force agreement; criteria cannot be combined across different attempts. For shelf-box pushing, a later force-regulation attempt remains valid if the box has already reached ad remains at the designated wall. Task-only SR counts task completion anywhere during active VLA control, independent of the force criteria.

Aggregation and tracking diagnostics. Prompt-, task-, and overall-level SRs use 10, 30, and 90 episodes per variant, respectively, with equally weighting across task–prompt combinations. Early-terminated episodes remain in the denominator; an episode is counted as successful if the success criterion was satisfied before termination. Table VII reports the complete prompt-level and gate-specific results.

TABLE VII

DETAILED CLOSED-LOOP VLA RESULTS (%). THE THREE TASK BLOCKS REPORT JOINT SUCCESS BY FORCE PROMPT AND TASK AVERAGE. THE OVERALL COLUMNS SHOW TASK COMPLETION, TASK PLUS PROMPT-BAND AGREEMENT, AND JOINT SUCCESS.

	Box pickup				Shelf pushing				Surface wiping				Overall		
Variant	G	F	S	Avg.	G	F	S	Avg.	G	F	S	Avg.	Task	+ band	Joint
C0	100.0	0.0	0.0	33.3	30.0	60.0	70.0	53.3	100.0	40.0	0.0	46.7	85.6	45.6	44.4
C1	80.0	20.0	0.0	33.3	80.0	90.0	90.0	86.7	90.0	90.0	10.0	63.3	81.1	61.1	61.1
C2	70.0	80.0	0.0	50.0	30.0	50.0	60.0	46.7	80.0	70.0	0.0	50.0	90.0	51.1	48.9
C3	90.0	90.0	0.0	60.0	70.0	80.0	70.0	73.3	100.0	100.0	0.0	66.7	86.7	72.2	66.7
C4	20.0	0.0	0.0	6.7	0.0	0.0	20.0	6.7	100.0	0.0	0.0	33.3	73.3	15.6	15.6
C5	40.0	0.0	0.0	13.3	50.0	80.0	70.0	66.7	100.0	0.0	0.0	33.3	71.1	38.9	37.8
C6	70.0	80.0	10.0	53.3	0.0	10.0	30.0	13.3	100.0	100.0	100.0	100.0	88.9	65.6	55.6
C7	80.0	90.0	60.0	76.7	70.0	70.0	70.0	70.0	100.0	100.0	100.0	100.0	94.4	88.9	82.2
C9	80.0	80.0	70.0	76.7	10.0	40.0	20.0	23.3	100.0	100.0	100.0	100.0	93.3	83.3	66.7

G/F/S: gentle/firm/strong; $n = 10$ per prompt, 30 per task, and 90 overall. Underlines mark the best result in each column, including ties. Variant settings are given in Table III.

For the default Opt2VLA configuration (C7), we additionally analyze issued force, measured force, and end-effector tracking error. Each episode contributes its mean over all command-active intervals; bars and whiskers show the mean and sample standard deviation across ten scenes for each prompt. These diagnostics include all episodes regardless of task success or measured contact, and therefore differ from the dual-active phases used for the mean-force-agreement criterion.

End-effector error is defined as the Euclidean distance between the VLA-issued controller target and the measured hand-body origin in the base-relative reset-heading frame. The instantaneous error is averaged within each episode before aggregation. These statistics describe command–execution discrepancies without uniquely attributing them to either the VLA policy or the low-level controller.

C7 tracking analysis. Figure 11 shows the force and end-effector tracking for the default Opt2VLA configuration. The largest difference between issued and measured mean force is 2.85 N for strong shelf-box pushing with the right hand, where the measured force exceeds the issued mcommand. Across prompt/hand groups, mean end-effector errors span 3.22–4.66 cm for box pickup, 7.19–7.52 cm for shelf-box pushing, and 5.48–6.31 cm for surface wiping. These statistics are computed over command-active windows, including unsuccessful episodes; the joint-SR mean-force-agreement criterion instead uses dual-active phases within individual attempts.

C7 failure breakdown. The Sankey diagram partitions each task's 30 episodes according to the first unmet requirement: task completion, prompt-band agreement, or mean-force agreement. All attempts are considered, but all three criteria must be satisfied within the same attempt. Missing command-active windows or dual-active phases remain failures. Ribbon widths represent episode counts rather than timesteps. These ordered gate failures describe the observed failure conditions but do not uniquely attribute them to either the VLA policy or the low-level controller.

For box pickup, 5 episodes fail task completion and 2 additionl episodes fail prompt-band agreement. For shelf-box pushing, all episodes meet the task-only criterion; 3 fail prompt-band agreement and 6 additional episodes fail mean-force agreement. Surface wiping passes all three criteria in all 30 episodes.

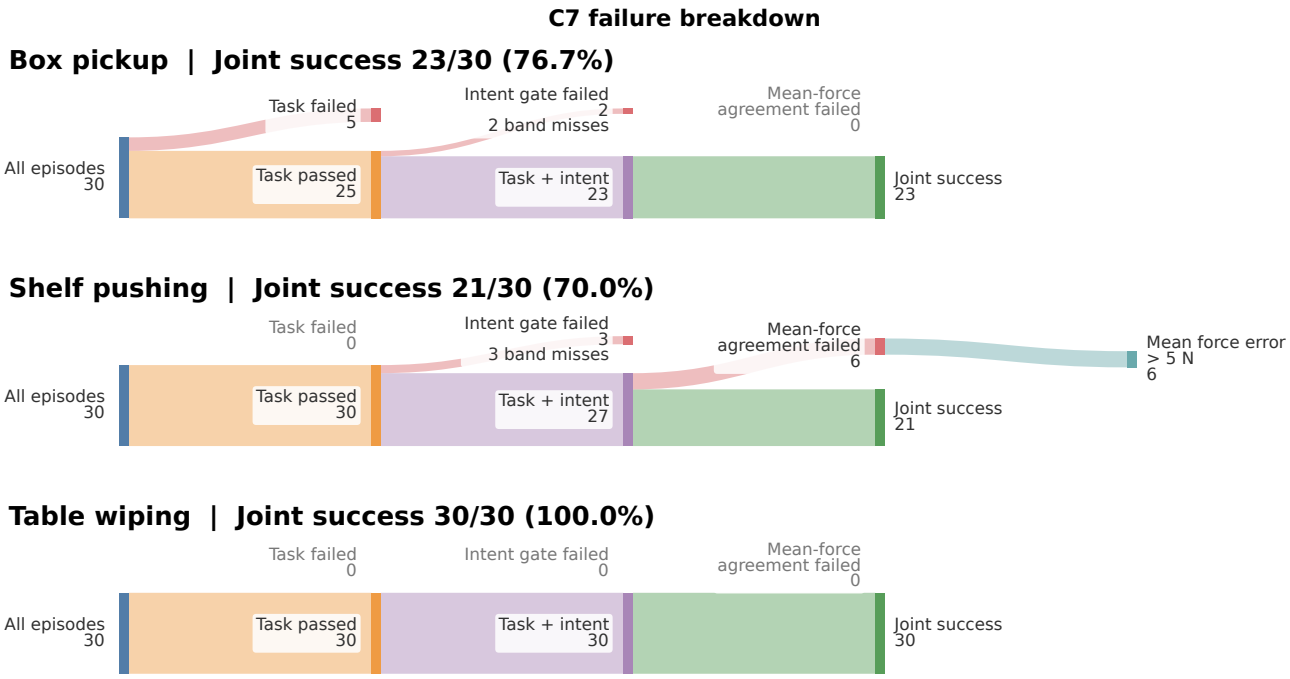

Fig. 10. **C7 failure breakdown.** Each task includes 30 episodes, partitioned by the first unmet criterion. Joint success requires task completion, prompt-band agreement, and mean-force agreement within the same attempt.

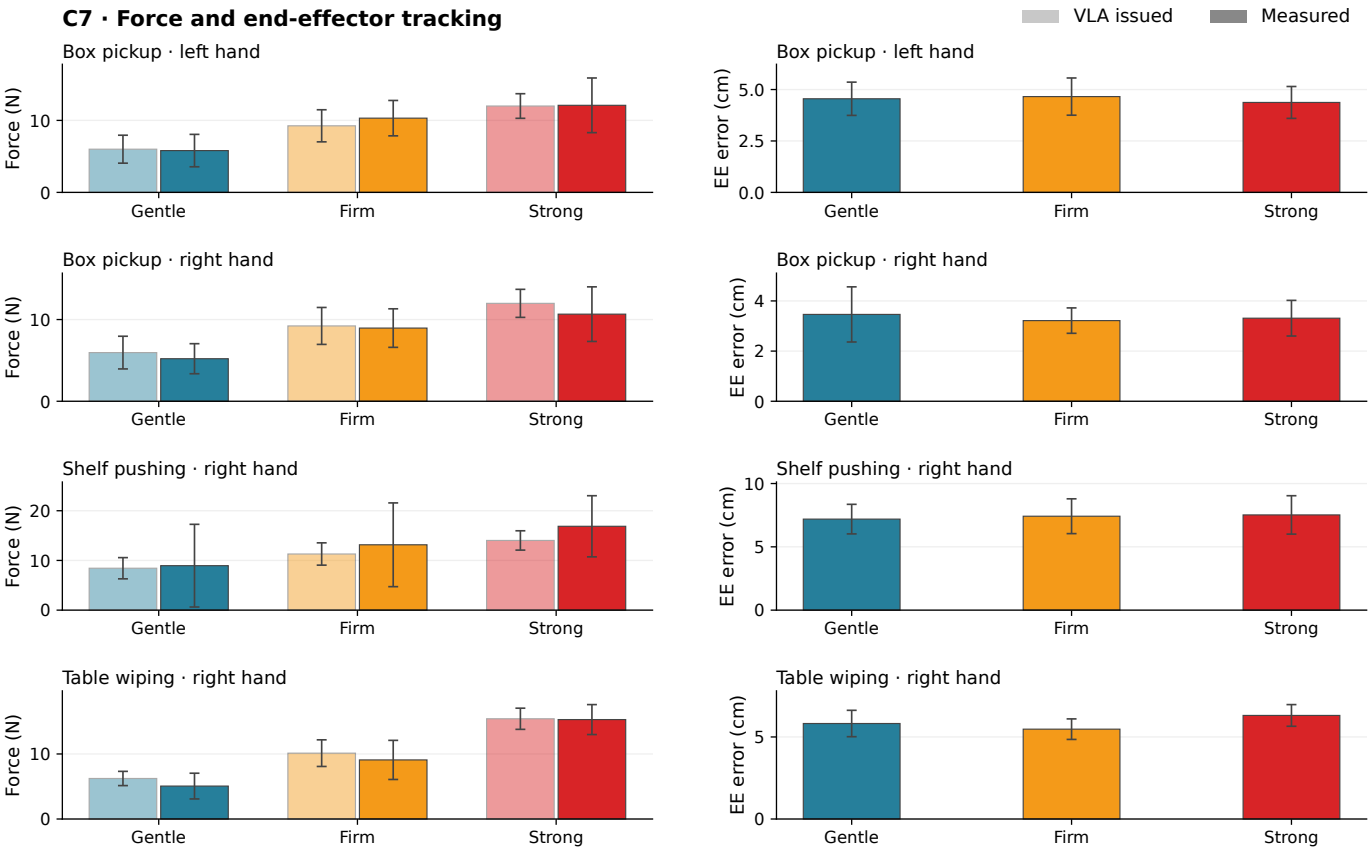

Fig. 11. **C7 tracking diagnostics.** Issued and measured force (left) and end-effector position error (right). Bars show means of episode means; whiskers show one sample standard deviation across ten scenes per prompt. All command-active intervals are included, regardless of task success or measured contact.