

Efficient and Compliant Control Framework for Versatile Human-Humanoid Collaborative Transportation

Shubham S. Kumbhar¹, Abhijeet M. Kulkarni¹, Panagiotis Artemiadis^{1*}

¹Mechanical Engineering Department, University of Delaware, 126 Spencer Lab, Newark, 19716, Delaware, United States of America.

*Corresponding author(s). E-mail(s): partem@udel.edu;
Contributing authors: shubhamk@udel.edu; amkulk@udel.edu;

Abstract

We present a control framework that enables humanoid robots to perform collaborative transportation tasks with a human partner. The framework supports both translational and rotational motions, which are fundamental to co-transport scenarios. It comprises three components: a high-level planner, a low-level controller, and a stiffness modulation mechanism. At the planning level, we introduce the Interaction Linear Inverted Pendulum (I-LIP), which, combined with an admittance model and an MPC formulation, generates dynamically feasible footstep plans. These are executed by a QP-based whole-body controller that accounts for the coupled humanoid-object dynamics. Stiffness modulation regulates robot-object interaction, ensuring convergence to the desired relative configuration defined by the distance between the object and the robot’s center of mass. We validate the effectiveness of the framework through real-world experiments conducted on the Digit humanoid platform. To quantify collaboration quality, we propose an efficiency metric that captures both task performance and inter-agent coordination. We show that this metric highlights the role of compliance in collaborative tasks and offers insights into desirable trajectory characteristics across both high- and low-level control layers. Finally, we showcase experimental results on collaborative behaviors, including translation, turning, and combined motions such as semi-circular trajectories, representative of naturally occurring co-transportation tasks.

Keywords: pHRI, Model Predictive Control, Admittance, I-LIP, Quadratic programs, Efficiency

1 Introduction

Humans routinely collaborate to accomplish physically demanding tasks, such as moving furniture across a room, carrying packages in a warehouse, or assisting elderly individuals with mobility. These activities underscore the importance of coordination, adaptability, and collaborative effort in achieving common objectives. Extending such collaboration to robots has the potential to transform multiple domains, from logistics

and manufacturing to healthcare and home assistance, by enabling humans and robots to work together as integrated partners. Over the past two decades, significant research has been conducted in the field of *physical human-robot interaction* (pHRI), as comprehensively reviewed by [Jarrassé et al. \(2012\)](#); [Lasota et al. \(2017\)](#); [Ajoudani et al. \(2018\)](#); [Vianello et al. \(2021\)](#).

Physical human-robot interaction (pHRI) describes scenarios in which a human and a robot physically interact to achieve a shared objective.

These interactions may be *direct*, such as guiding a robot by the hand (Lefèvre et al. 2024), or *indirect*, such as jointly manipulating an object (Agravante et al. 2019). In both cases, the partners must regulate their interaction forces to accomplish task-specific objectives. Effective pHRI requires the robot to remain dynamically stable under external loads, adapt in real-time to unknown human intent, and maintain task efficiency and safety in uncertain environments. These requirements make pHRI inherently more challenging than conventional robotic control, as the robot must both predict motion and execute whole-body actions while accounting for the presence of a collaborator.

1.1 Previous Works

Robotic manipulators have been extensively employed in pHRI for tasks such as handovers (Huber et al. 2008; Edsinger and Kemp 2007), carrying unknown loads (Mujica et al. 2023), and collaborative lifting (DelPreto and Rus 2019). For instance, DelPreto and Rus (2019) leveraged human arm muscle activity to infer the desired lifting height of an object during cooperative tasks. Similarly, Rozo et al. (2014) demonstrated that collaborative skills can be transferred from human-human demonstrations to robotic manipulators. These studies demonstrate the effectiveness of robotic manipulators in physical interaction tasks, but their fixed base restricts the reachable workspace and prevents the mobility needed for tasks such as collaboratively transporting objects over long distances.

To address the limitations of fixed-base manipulators, mobile manipulator setups have been proposed by Kosuge et al. (2000, 2001); Suda and Kosuge (2002). These works developed control strategies that regulate the compliance between the mobile base and the object being carried. More recently, Xing et al. (2021) introduced a control framework based on skill transfer for collaborative lifting and carrying tasks using wheeled mobile manipulators. Beyond object transportation, a compliant control strategy has been proposed in assisted walking (Li et al. 2024). While these robots provide greater maneuverability and extend the operational workspace, these systems are fundamentally optimized for flat, structured surfaces (Bruzzzone and Quaglia 2012). In contrast,

human environments often contain irregular terrain, stairs, and cluttered layouts, where wheeled platforms are less effective (Bruzzzone and Quaglia 2012). This gap highlights the need for robots with legged bases, whose inherent versatility and maneuverability make them better suited for collaboration in human-centric environments.

The efficiency and fluidity observed in human-human collaboration provide a strong motivation for extending similar principles to humanoid robots, giving rise to the field of *physical human-humanoid interaction* (pHHI). Unlike wheeled mobile manipulators, humanoids can naturally navigate uneven terrains (Radosavovic et al. 2024), climb stairs (Caron et al. 2019), and operate within cluttered or constrained spaces (Narkhede et al. 2022; Kumbhar et al. 2025), making them particularly well-suited for tasks in warehouses, homes, and healthcare settings.

Beyond functional capabilities, humanoids have attracted increasing attention in both industry and academia due to their dexterity and versatility in performing a wide range of tasks. Recent market reports even project widespread global deployment of humanoids by 2040 (Neumann 2025), underscoring their anticipated impact across multiple sectors. However, the task of pHHI is considerably more challenging. Unlike wheeled bases, which are mechanically simpler and dynamically more stable, the task of maintaining balance, stability, and robustness for legged humanoids is much more challenging. These challenges necessitate advanced control frameworks that can ensure safe, stable, and compliant interaction between humanoids and human partners.

One of the earliest demonstrations of pHHI was presented by Yokoyama et al. (2003), where the HRP-2P humanoid robot was used to collaboratively move a panel with a human operator. Agravante et al. (2014) implemented a control architecture for humanoid collaboration based on a Stack of Tasks formulation. Their approach provided a vision algorithm to estimate the desired object states. These desired states were subsequently used for admittance control of the humanoid. A different approach was adopted by Motahar et al. (2015, 2017), where interaction forces were leveraged for adaptation of the walking limit cycle, while the arms were impedance

controlled to ensure safe physical interaction. Subsequent works have explored the use of intent-carrying forces to enable switching, either among a library of limit cycles (Veer et al. 2017) or among Dynamic Motion Primitives (DMPs) (Chand et al. 2022).

Current research on humanoid walking with optimization-based methods typically formulates the control hierarchy at two levels. At the high level, center of mass (CoM) trajectories and footstep locations are planned, while at the low level, these plans are realized on the physical robot through whole-body control. Model Predictive Control (MPC) has emerged as a widely used approach for high-level planning, leveraging simplified dynamic models such as the Linear Inverted Pendulum (LIP) (Narkhede et al. 2023; Kajita et al. 2001), and the Single Rigid Body (SRB) (Ding et al. 2022). At the low level, Quadratic Programming (QP) formulations are employed to compute feasible joint torques and contact forces that ensure stability and adherence to the planned trajectories (Nguyen and Sreenath 2015). Current work in pHHI, including our own, adopts a similar hierarchical structure, employing an MPC-based high-level planner for footstep generation, coupled with a QP-based low-level controller to track and execute the motion plan on the Digit humanoid robot (Agravante et al. 2019; Kumbhar and Artemiadis 2025).

The work by Agravante et al. (2016) introduced a hierarchical MPC-QP framework specifically for pHHI. The control for the humanoid as a follower incorporated an impedance model of the CoM directly into the MPC cost function to regulate interaction behavior, and footstep patterns were generated accordingly. However, they used a quasi-static motion assumption to transform the external wrench from the sensor frame to the CoM frame, and it was also assumed to remain constant over the MPC horizon. Furthermore, the impedance model specifies a desired impedance between the robot CoM and the origin, implying that the human partner needed to exert significant effort to displace the robot from its starting configuration. Later, Agravante et al. (2019) tracked these high-level plans via a QP that explicitly considered external forces. While knowledge of these forces enabled compensatory actions for stability and trajectory tracking, the robot did not actively regulate them.

Kumbhar and Artemiadis (2025) addressed these challenges by coupling an object-informed QP for low-level control with an admittance model designed to regulate the relative motion between the humanoid’s CoM and the object. Crucially, the stiffness modulation strategy adopted in their approach utilized desired states obtained by integrating the admittance model at each control iteration, starting from the previous desired state. We characterize this approach as *global compliance*, as it effectively forces the system to track a cumulative, ideal trajectory generated from the start of the task. Consequently, this formulation prevents *local compliance*, which we define as deriving the desired response relative to the robot’s current actual state. By prioritizing global tracking, the controller behaves rigidly in the face of transient deviations, failing to provide the intended compliant response to perturbations. Furthermore, the scope of their study was limited to straight-line motion, leaving the turning maneuvers essential for collaborative transportation unaddressed.

In pHHI, although it is important for the humanoid to be stable while assisting with the task, the influence of the control framework on human performance and task-level efficiency has often not been addressed (Kumbhar and Artemiadis 2025; Agravante et al. 2019). Rapetti et al. (2023) proposed a control framework for human-humanoid ergonomic lifting that accounts for joint-level efficiency and incorporates a detailed model of the human. While such approaches provide valuable insights into human biomechanics, collaborative transportation also requires evaluating task-level factors such as synchronization between partners and the overall cost of co-transportation.

1.2 Main Contributions

This paper presents a comprehensive framework for enhancing the compliant behavior of humanoid robots when physically collaborating with humans to transfer objects, with a particular focus on scenarios involving the maneuvering of these objects. We extend the work done by Kumbhar and Artemiadis (2025), to propose a three-level hierarchical control architecture that integrates: (i) an admittance model to regulate the desired interaction behavior between the humanoid and the object, including turning maneuvers arising from

object rotation; (ii) a high-level planner based on the *Interaction Linear Inverted Pendulum* (I-LIP), a **novel** simplified model incorporating variable-stiffness springs and dampers between the centers of mass (CoMs) of the robot and the object, formulated within an MPC framework; and (iii) a low-level QP that explicitly incorporates object dynamics to regulate interaction forces and control the object’s position and orientation. The use of coupled dynamics has been shown to be effective, as demonstrated by Rapetti et al. (2023, 2021).

A key feature of our framework is that compliance arises from both footstep planning and hand-object interactions. While the MPC generates foot placements that encode compliant responses to perturbations, additional objectives on the CoMs of the robot and the object in the QP ensure that the hands contribute actively to compliance. To ensure convergence to a desired distance between the object and the humanoid, we implement a gain modulation strategy for adjusting the spring constant of the I-LIP. These features promote *local* compliance properties to the humanoid.

Additionally, we conduct an extensive analysis of the resulting framework to evaluate the role of compliance in enhancing task and human efficiency during collaborative object transportation while preserving robot stability. In particular, we introduce a quantitative efficiency metric to show that a low-compliance objective for footstep planning and a high-compliance objective for low-level control improve the efficiency of collaboration, while simultaneously preserving humanoid stability. Another distinguishing aspect of this work is its focus on carrying objects of substantial weight. The experimental validation considers a 15 kg box, representative of loads that typically require two humans to co-manipulate, a scenario not addressed in existing literature.

We validate the proposed framework on the Digit humanoid (Agility Robotics) through real-world experiments. Our results demonstrate that the framework enables compliant human–humanoid collaboration during complex maneuvers involving both translation and turning, while maintaining robustness to external perturbations and variations in human motion.

The remainder of this paper is organized as follows. Section 2 presents the proposed hierarchical control framework, including the I-LIP

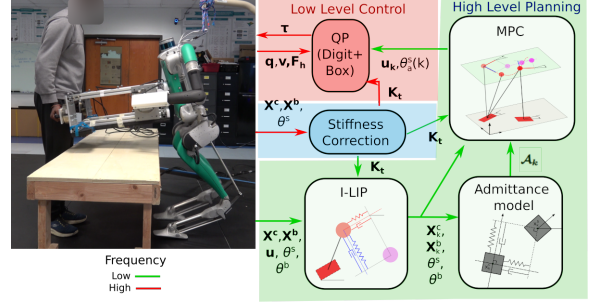


Fig. 1 Proposed control framework (right) for collaborative transportation tasks between a human and a humanoid robot (left).

model, admittance-based compliance formulation, MPC-based walking pattern generator, WBC formulation, and stiffness modulation strategy. Section 3 describes the experimental setup and presents the evaluation of the proposed framework through real-world human–humanoid co-transportation experiments. Finally, Section 4 concludes the paper and discusses future directions.

2 Control Architecture for Collaborative Transportation

We address collaborative transportation tasks in which the Digit humanoid and a human jointly transfer an object (box), as illustrated in Fig. 1. In this setting, the human assumes the role of the leader, actively maneuvering the object according to their intent, while the humanoid functions as a follower, passively adapting its motion to the object’s movement. Building upon the work presented by Kumbhar and Artemiadis (2025), we extend the follower’s capabilities to include forward, lateral, and rotational maneuvers, thereby providing the human collaborator greater freedom in executing the co-transportation task.

Figure 1(right) illustrates the proposed control framework for physical human–humanoid collaboration. The I-LIP and the admittance model, together with the MPC, form the high-level planning layer, which generates footstep placements and CoM trajectories. The I-LIP (Fig. 2(a)) serves as a simplified predictive model for planning purposes, while the admittance model (Fig. 2(b)) dictates the desired compliant behavior between the humanoid and the object. Varying admittance gains enables behaviors ranging from highly

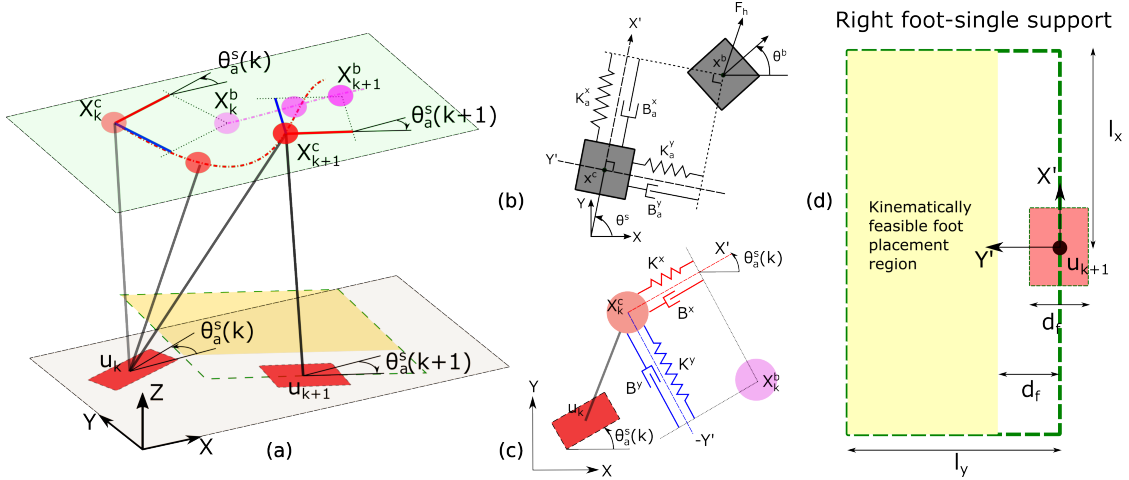


Fig. 2 (a) I-LIP dynamics during the k^{th} walking step with the left leg as the stance foot, denoted by $\mathbf{u}_k$. The red and blue solid lines represent spring-damper elements aligned along the local X' and Y' axes, respectively. The red dashed trajectory indicates the evolution of the robot's center of mass (CoM), while the purple dashed trajectory corresponds to the object's CoM. Temporal progression is illustrated through decreasing transparency, where increased opacity represents forward movement in time. The red rectangle denotes the stance foot position, and the yellow rectangle indicates the admissible region for the next foot placement based on kinematic feasibility. (b) Admittance-based interaction model, abstracted as two point masses, representing the CoMs of the robot and the object interconnected by spring damper pairs in the X' and Y' directions. This model captures the compliant coupling between the robot and the object and enables responsive interaction. (c) Top-down view of the I-LIP configuration at the beginning of the k^{th} step, highlighting the orientation of spring-damper elements in the horizontal X' - Y' plane. (d) Admissible foot placement region with the right leg as the stance foot. The yellow region defines the kinematically feasible area for placing the left foot in the next step. This region is constrained by the robot's reachability and ensures both feasibility and stability in step planning.

compliant motion to more assertive tracking. The MPC integrates the I-LIP dynamics with the desired states provided by the admittance model to compute optimal footstep locations that satisfy the modeling constraints.

The resulting high-level plans are executed on the robot by the WBC, formulated as a QP. Unlike conventional humanoid control approaches that consider only the robot's own dynamics, our WBC models the coupled humanoid-object system, enabling direct regulation of interaction forces. The stiffness modulation module adjusts the parameters of the I-LIP's virtual coupling in real time to improve compliance tracking and maintain a desired relative motion between the humanoid and the object.

Forward and lateral motions are defined as movements of the humanoid along its local X' and Y' axes, respectively, as illustrated in Fig. 2(c), while turning motion corresponds to rotations about the Z -axis. For clarity, both the local coordinate frame and the global (world) frame are shown in the figure.

Let $\mathbf{x}^c = [x^c, y^c]^\top$ and $\dot{\mathbf{x}}^c = [\dot{x}^c, \dot{y}^c]^\top$ denote the global X - and Y -component position and

velocity, respectively, of the Digit's center of mass (CoM). The global position of the stance foot is given by $\mathbf{u} = [u^x, u^y]^\top$, with the Z -component of all foot strikes assumed to be zero. Similarly, the object's global position and velocity are denoted by $\mathbf{x}^b = [x^b, y^b]^\top$ and $\dot{\mathbf{x}}^b = [\dot{x}^b, \dot{y}^b]^\top$, respectively.

For compactness, we define the concatenated state vectors $\mathbf{X}^c = [\mathbf{x}^c, \dot{\mathbf{x}}^c]^\top$ for the humanoid and $\mathbf{X}^b = [\mathbf{x}^b, \dot{\mathbf{x}}^b]^\top$ for the object. The yaw angles of the stance foot and the object are denoted by θ^s and θ^b , respectively. As shown in Fig. 1, the high-level planning block takes the tuple $(\mathbf{X}^c, \mathbf{X}^b, \mathbf{u}, \theta^s, \theta^b)$ as input for subsequent computations in the I-LIP model, the admittance model, and the MPC module, described in the following subsections.

2.1 Interaction Linear Inverted Pendulum (I-LIP)

Hierarchical humanoid control architectures commonly employ simplified dynamic models for CoM motion and footstep planning, while whole-body controllers execute the resulting trajectories on

the full-order system (Wensing et al. 2024). Simplified models provide computationally tractable planning by reducing the dimensionality and non-linearity of the optimization problem. In this work, we adopt the I-LIP model for prediction and motion planning.

Figure 2(a) illustrates the I-LIP model used for predicting the evolution of the robot’s CoM. The I-LIP extends the classical LIP model (Kajita et al. 2001) to incorporate the influence of interaction forces and object motion into footstep and CoM planning. The I-LIP consists of a massless prismatic leg and two point masses representing the CoMs of the robot and the object. One end of the leg pivots at the stance foot contact point on the ground, while the other is attached to the point mass corresponding to the robot’s CoM. The two point masses are connected by linear springs and dampers in both the X' and Y' directions, as shown in Fig. 2(c). The stance foot also has a yaw (heading) angle, and the springs and dampers are oriented along the local coordinate axes defined by this heading. In the model, the object is assumed to move with a constant velocity, denoted by $\mathbf{v}_d^b = [v_d^{bx}, v_d^{by}]^\top$. This velocity is estimated using an exponential moving average of the current measured object velocity, computed as:

$$\mathbf{v}_d^b(t) = \alpha \mathbf{v}_d^b(t - dt) + (1 - \alpha) \dot{\mathbf{x}}^b \quad (1)$$

where $\alpha \in (0, 1]$ is the smoothing factor, t is the current time, and $dt = 2$ ms is the time period of the low-level control loop. The k^{th} step of the I-LIP occurs over the time interval $[kT, (k+1)T)$ between two consecutive foot strikes, where T denotes the step duration. In Fig. 2(a,c), $\mathbf{u}_k = \mathbf{u}(t = kT)$ and θ_k^s represent the position and the heading (yaw) of the stance foot during the k^{th} step. The vectors $\mathbf{X}_k^c = \mathbf{X}^c(t = kT)$ and $\mathbf{X}_k^b = \mathbf{X}^b(t = kT)$ denote the CoM state vectors of the robot ($\mathbf{X}^c$) and the object ($\mathbf{X}^b$), respectively, at the start of the k^{th} step. The stepping sequence alternates between the left and right foot; thus, if $\mathbf{u}_k$ corresponds to the left stance foot, then $\mathbf{u}_{k+1}$ corresponds to the right stance foot.

The dynamics of the CoM of the robot in I-LIP are described by

$$\begin{aligned} m_c \ddot{\mathbf{x}}^c &= m_c \omega_o^2 (\mathbf{x}^c - \mathbf{u}) + \\ &\mathbf{R}(\theta^s) \mathbf{K}_t (\mathbf{R}(\theta^s))^\top (\mathbf{x}_{\text{init}}^b + \mathbf{v}_d^b t - \mathbf{x}^c - \mathbf{x}_m^d) \\ &+ \mathbf{R}(\theta^s) \mathbf{B} (\mathbf{R}(\theta^s))^\top (\mathbf{v}_d^b - \dot{\mathbf{x}}^c) \end{aligned} \quad (2)$$

where $\omega_o = \sqrt{g/h}$, g is the magnitude of the gravitational acceleration, h is the height of the robot’s CoM from the ground, $\mathbf{x}_{\text{init}}^b$ is the initial position of the CoM of the object, and $\mathbf{x}_m^d$ is the x-y component natural length of the spring. $\mathbf{K}_t = \text{diag}(K_t^x, K_t^y)$ is the spring constant matrix at the instant the high-level planner is initiated. The subscript t in $\mathbf{K}_t$ and K_t^x indicates that these values vary with time according to the stiffness modulation block in Fig. 1. However, within the high-level planning block they are fixed to the value at the instant the planner is invoked and remain constant until the planning process is completed. $\mathbf{B} = \text{diag}(B^x, B^y)$ is the damping constant matrix. The purpose of the springs and dampers in the I-LIP model is not to estimate the interaction forces acting on the CoM. Instead, they are introduced to induce corrective adjustments in the footstep planning to compensate for interaction forces and moments arising from the heavy load. These corrections are further regulated through stiffness modulation (Section 2.5), ensuring that the desired task objectives, such as target distance and velocity, are ultimately achieved. By applying the analytical solutions of the non-homogeneous linear differential equations in (2), the step-to-step discrete dynamics of the I-LIP f_c can be derived as shown below.

$$\mathbf{X}_{k+1}^c = f_c(\mathbf{X}_k^c, \mathbf{x}_k^b, \mathbf{v}_d^b, \mathbf{u}_k, \theta_k^s, \mathbf{K}_t, T) \quad (3)$$

Similarly, the step-to-step evolution of the CoM of the object under the constant object velocity assumption is given by:

$$\mathbf{X}_{k+1}^b = f_b(\mathbf{x}_k^b, \mathbf{v}_d^b, T) = \begin{bmatrix} \mathbf{x}_k^b + \mathbf{v}_d^b T \\ \mathbf{v}_d^b \end{bmatrix} \quad (4)$$

If we assume that the system is in the $(k-1)^{\text{th}}$ step, the I-LIP block in Fig. 1 utilizes the functions f_c and f_b , and $\mathbf{x}_{\text{init}}^b$ as the current state to predict the COM states of the robot ($\mathbf{X}_k^c$) and the object ($\mathbf{X}_k^b$) at the end of the current step. As illustrated in Fig. 1, the tuple $(\mathbf{X}_k^c, \mathbf{X}_k^b, \theta^s, \theta^b)$

is subsequently provided to the admittance model and MPC blocks for further planning.

Due to the physical limitations of the robot, the I-LIP formulation additionally incorporates kinematic constraints on the admissible region of the subsequent foot placement. Figure 2(d) illustrates the feasible foot placement region for the subsequent step of the I-LIP. This region is parameterized by the following quantities: l_x : maximum footstep length in the forward direction, l_y : maximum footstep length in the lateral direction, and d_f : foot width. The feasible foot placement region is defined as:

$$\begin{aligned} \mathcal{U}_k(\theta_k^s) = \{ & (u^x, u^y) | \\ & -l_x \leq (\mathbf{R}_1(\theta_k^s))^\top (u_{k+1}^x - u_k^x) \leq l_x, \\ & d_f \leq n (\mathbf{R}_2(\theta_k^s))^\top (u_{k+1}^y - u_k^y) \leq l_y \}, \end{aligned} \quad (5)$$

where $n = 1$ if the stance foot is the right foot, and $n = -1$ otherwise. The subscripts 1 and 2 in $(\mathbf{R}(\cdot))^\top$ denote the first and second rows of the rotation matrix, respectively. For the lateral Y' direction, an additional constraint ensures that the left and right legs do not cross. Furthermore, a clearance equal to the foot width is included to prevent collisions between the swing foot and the stance foot.

The states predicted at the end of the step are subsequently used in the admittance model to compute the desired CoM motion corresponding to a specified compliant behavior. These results are then provided to the MPC module for generating the footstep pattern.

2.2 Compliance Shaping via Admittance Model

Admittance control is a widely used technique in human-robot collaboration (Ajoudani et al. 2018) for enabling safe interaction and achieving compliant behavior of the robot during physical contact. In this work, an admittance model is integrated with the footstep pattern generator to regulate the compliant behavior of the humanoid during co-transportation. The objective is to realize compliant behavior by leveraging both the stepping pattern and the arm motions within the whole-body control framework.

The admittance model consists of two point masses connected by a spring and a damper, as

illustrated in Fig. 2(b). The robot and the object are represented as point masses with masses m_c and m_b , respectively. These are coupled through a pair of springs and dampers oriented along the heading direction defined by the stance foot yaw.

The stiffness and damping parameters of these elements can be tuned to achieve the desired compliant behavior between the robot and the object. For instance, a lower stiffness value increases the robot's compliance relative to the object, whereas a higher stiffness reduces compliance.

We assume that the object moves with a constant velocity $\mathbf{v}_d^b$, as computed in the previous subsection. This constant-velocity assumption in the I-LIP and admittance model applies only to future box velocity predictions over the 1.5 s MPC horizon. However, $\mathbf{v}_d^b$ is updated at each control iteration (500Hz) via (1). Consequently, the humanoid is able to handle non-constant velocity motions, as demonstrated in Section 3.2. Under this assumption, the dynamics of the robot's CoM in the admittance model can be expressed as:

$$\begin{aligned} m_c \ddot{\mathbf{x}}^c = & \mathbf{R}(\theta^s) \mathbf{K}_a (\mathbf{R}(\theta^s))^\top (\mathbf{x}^b - \mathbf{x}^c - \mathbf{x}^d) \\ & + \mathbf{R}(\theta^s) \mathbf{B}_a (\mathbf{R}(\theta^s))^\top (\mathbf{v}_d^b - \dot{\mathbf{x}}^c) \end{aligned} \quad (6)$$

$$\mathbf{x}^b = \mathbf{x}_{\text{init}}^b + \mathbf{v}_d^b t \quad (7)$$

Here, $\mathbf{x}_{\text{init}}^b$ is the initial position of the CoM of the object. $\mathbf{K}_a = \text{diag}(K_a^x, K_a^y)$ and $\mathbf{B}_a = \text{diag}(B_a^x, B_a^y)$ are positive definite diagonal matrices. It is important to note here that dynamics in (2) and (6) might look similar, but they are completely different models for their respective purposes. While the purpose of I-LIP dynamics in (2) is to aid footstep planning through the MPC, the objective of the admittance model is to provide the desired CoM states and footstep orientations of the robot at the end of each future step. For a prediction horizon of N steps, the admittance model dynamics are integrated to obtain the N predicted end-of-step COM states and corresponding footstep orientations.

Let the desired orientation of the stance foot be denoted by θ_a^s . Similar to previous definitions, $\theta_a^s(k)$ denotes the desired orientation of the stance foot during the k^{th} step. The orientations are computed using a rotational spring-mass admittance

model, expressed as:

$$\ddot{\theta}_a^s = k_P^\theta (\theta_d^b - \theta_a^s) - k_D^\theta \dot{\theta}_a^s \quad (8)$$

Here, k_P^θ and k_D^θ are positive constants that rule the compliant nature of the stance foot yaw with respect to the object yaw. A damping term ($k_D^\theta \dot{\theta}_a^s$) is added to stabilize the response. θ_d^b denotes the estimated intended yaw of the object. This estimate is computed in a manner similar to $\mathbf{v}_d^b$, as

$$\theta_d^b(t) = \beta \theta_d^b(t - dt) + (1 - \beta) \theta^b \quad (9)$$

where $\beta \in (0, 1]$ is the smoothing factor. The dynamics for θ_a^s (8) are derived under the assumption that θ_d^b remains constant over the next N steps. These dynamics are integrated from the initial value θ_{k-1}^s to obtain $\theta_a^s(k+i) \forall i = 0, 1, 2, \dots, N-1$. These desired orientations are further utilized in the admittance model to obtain the desired CoM states. It is important to note that, within each step, the stance foot orientation is fixed at θ_k^s . Consequently, the actual evolution of the heading follows a stepwise profile rather than the smooth trajectories given by (8). Nonetheless, these profiles coincide periodically at the end of each step.

Using the analytical solution of the non-homogeneous differential equation (6), along with the desired stance foot orientations obtained above, we define a function f_a that maps the robot's CoM state at the start of the k^{th} step, $(\mathbf{x}_k^c, \dot{\mathbf{x}}_k^c)$, to the CoM state at the start of the $(k+1)^{\text{th}}$ step, $(\mathbf{x}_{k+1}^c, \dot{\mathbf{x}}_{k+1}^c)$:

$$\mathbf{x}_{k+1}^c, \dot{\mathbf{x}}_{k+1}^c = f_a(\mathbf{x}_k^c, \dot{\mathbf{x}}_k^c, \mathbf{x}_k^b, \mathbf{v}_d^b, \theta_a^s(k), T). \quad (10)$$

Using $\mathbf{X}_k^c$ and $\mathbf{X}_k^b$ as inputs (as shown in Fig. 1), and applying (10), the admittance model block computes $\mathbf{x}_a^c(k+i) = \mathbf{x}_{k+i}^c$, $\dot{\mathbf{x}}_a^c(k+i) = \dot{\mathbf{x}}_{k+i}^c$, $\forall i = 1, 2, \dots, N$, where N denotes the prediction horizon of the MPC, i.e., the number of future steps considered.

Let the output of the admittance model block be denoted by the set $\mathcal{A}_k$, where:

$$\mathcal{A}_k = \bigcup_{i=1}^N \{\mathbf{x}_a^c(k+i), \theta_a^s(k)\} \quad (11)$$

This set is passed on as goal positions in the MPC formulation that uses I-LIP and computes desired footstep locations.

2.3 Walking Pattern Generation using MPC

MPC is extensively used in the literature for footstep planning of legged robots (Katayama et al. 2023). Its predictive decision-making capability, inherent constraint-handling, robustness, and ability to accommodate multi-objective optimization make it a strong candidate for this task.

We formulate the walking pattern generator for physical human-humanoid collaboration as an MPC problem. The MPC is formulated as a QP subject to constraints on foot placement and CoM states. It takes the current states ($\mathbf{X}_k^c, \mathbf{X}_k^b$), the goal states $\mathcal{A}_k$, and the stiffness parameters $\mathbf{K}_t$ as inputs and optimizes future foot placements of the I-LIP model over an N -step prediction horizon to track the desired CoM trajectory generated by the admittance model. The outputs is the optimal (x, y) position of the stance foot for the next step, $\mathbf{u}_k$. The formulation is as follows:

$$\underset{\mathbf{X}_{i+1}^c, \mathbf{u}_i}{\text{minimize}} \quad \sum_{i=k}^{k+N-1} J_i(\mathbf{x}_i^c, \mathbf{u}_{i-1}, \theta_a^s(i-1)) \quad (12)$$

$$\text{subject to} \quad \mathbf{X}_{i+1}^c = f_c(\mathbf{X}_i^c, \mathbf{x}_i^b, \mathbf{v}_d^b, \mathbf{u}_i, \theta_a^s(i), \mathbf{K}_t, T) \quad (13)$$

$$\mathbf{u}_i = (u_i^x, u_i^y) \in \mathcal{U}_i \quad (14)$$

The decision variables of the optimization are the robot's CoM states, $\mathbf{X}_{i+1}^c$, and the stance foot positions, $\mathbf{u}_i$, for all $i \in \{k, k+1, \dots, k+N-1\}$. It is important to note that the yaw angles are not included as decision variables; instead, the desired yaw values are provided directly as parameters of the optimization. This choice ensures that the MPC formulation remains convex. The optimization is subject to two constraints: (i) the step-to-step dynamics of the I-LIP model (13), and (ii) the foot placement constraint (14). The cost J_i is defined as a weighted sum of two individual cost terms, $\Phi_{i,1}$ and $\Phi_{i,2}$. These terms are defined identically for all values of i . For instance, for $i = k$, these are given by the expressions detailed below.

2.3.1 Goal position tracking

This objective term penalizes the deviation between the robot's CoM position and the corresponding goal position, thereby promoting footstep patterns that minimize this error. The cost is defined as:

$$\Phi_{k,1} = (\mathbf{x}_k^c - \mathbf{x}_a^c(k))^\top \mathbf{K}_{\theta_a^s(k-1)} (\mathbf{x}_k^c - \mathbf{x}_a^c(k)), \quad (15)$$

$$\mathbf{K}_{\theta_a^s(k-1)} = \mathbf{R}(\theta_a^s(k-1))^\top \mathbf{K}_\Phi \mathbf{R}(\theta_a^s(k-1)), \quad (16)$$

where $\mathbf{K}_\Phi$ is a positive-definite diagonal gain matrix. The gains are first selected in the local X' (forward) and Y' (lateral) directions, and then rotated into the world frame using the current stance foot yaw $\theta_a^s(k-1)$. This formulation enables independent tuning of stability in the forward and lateral directions.

2.3.2 CoM-foot distance regulation

The second cost term penalizes the distance between the robot's CoM and the stance foot at the end of the step. This discourages footstep placements that lead to increased distance between CoM and the foot, which can cause the robot to fall. A similar principle appears in the capture point concept for the LIP (Koolen et al. 2012), where a larger separation between the capture point and the foot corresponds to a state closer to a fall.

For our case, the cost is defined as:

$$\Phi_{k,2} = (\mathbf{x}_k^c - \mathbf{u}_{k-1})^\top \mathbf{B}_{\theta_a^s(k-1)} (\mathbf{x}_k^c - \mathbf{u}_{k-1}), \quad (17)$$

$$\mathbf{B}_{\theta_a^s(k-1)} = \mathbf{R}(\theta_a^s(k-1))^\top \mathbf{B}_\Phi \mathbf{R}(\theta_a^s(k-1)), \quad (18)$$

where $\mathbf{B}_\Phi$ is a positive-definite diagonal gain matrix.

The cost J_i in the MPC objective is then given by:

$$J_i(\mathbf{x}_i^c, \mathbf{u}_{i-1}) = \phi_1 \Phi_{i,1} + \phi_2 \Phi_{i,2},$$

where ϕ_1 and ϕ_2 are diagonal matrix weights chosen according to the relative importance of each term. Once the MPC problem is solved, the next planned footstep location $\mathbf{u}_k$ and orientation $\theta_a^s(k)$ are sent to the low-level controller for foot placement tracking, as shown in Fig. 1.

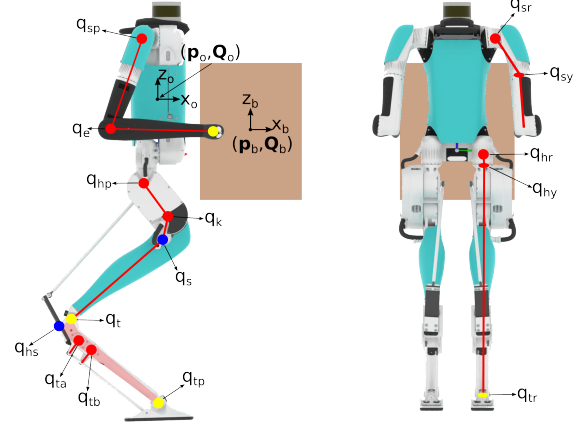


Fig. 3 Configuration space of the coupled system comprising of Digit and the box.

2.4 Whole Body Control using a Quadratic Program

The objective of the WBC, or the low-level control, is to execute the high-level plan generated by the MPC on Digit. We formulate the WBC as a QP subject to a set of constraints and objectives. The objectives encompass both task-space and joint-space goals, such as swing foot trajectory tracking, torso orientation regulation, and other tasks described in detail below. The constraints ensure that the resulting control commands are consistent with the robot's dynamics and comply with any additional bounds imposed on the states or decision variables.

As shown in Fig. 1, the QP receives as inputs the current state of the system $(\mathbf{q}, \mathbf{v})$, the forces applied by the human, and the footstep plan $(\mathbf{u}_k, \theta_a^s(k))$ provided by the high level planning block. The solution of the QP yields the joint torque commands $\boldsymbol{\tau}$, which are then sent to Digit for execution.

In the subsections that follow, we first provide a description of the humanoid robot Digit and define the system's state space $(\mathbf{q}, \mathbf{v})$. We then present the constraints and objectives used in the formulation of the low-level QP.

2.4.1 Coupled System: Digit and the Object

Digit is modeled as a tree structure composed of rigid bodies connected through joints. Its structure can be divided into five major assemblies:

torso, left arm, left leg, right arm, and right leg. The overall joint configuration is illustrated in Fig. 3. The torso, arms, and legs have 6, 4, and 8 degrees of freedom (DoFs), respectively. In total, Digit possesses 30 DoFs, of which 20 are actuated and 4 are passive; the remaining 6 correspond to the position and orientation of the torso.

Let the position of the torso with respect to the inertial frame be denoted by $\mathbf{p}_o = (x_o, y_o, z_o) \in \mathbb{R}^3$, and let $\mathbf{Q}_o \in SO(3)$ represent its orientation. The torso serves as the parent body for both the legs and arms.

Each leg, $i \in \{L, R\}$, is connected to the torso through a 3-DoF hip joint parameterized by yaw, roll, and pitch angles: q_{hy}^i , q_{hr}^i , and q_{hp}^i , respectively. This joint connects the torso to the thigh. At the distal end of the thigh, a knee joint (q_k^i) and a shin spring (q_s^i) connect the thigh to the shin link. The shin and the tarsus are linked via a joint parameterized by q_t . The heel-spring (q_{hs}^i) attached to the tarsus connects to the thigh to form a compliant kinematic loop. The leg ends with a heel with two degrees of freedom described by corresponding roll (q_{tr}^i) and pitch (q_{tp}^i) angles. It is actuated using two four-bar mechanisms parametrized by q_{ta} and q_{tb} . Finally, each arm, $i \in \{L, R\}$, attached to the torso is parametrized by shoulder yaw (q_{sy}^i), roll (q_{sr}^i), pitch (q_{sp}^i) and elbow (q_e^i) angles. Each leg/arm pair configuration is denoted by $\mathbf{q}_i$ as shown below.

$$\mathbf{q}_{i \in \{L, R\}} = (q_{hy}^i, q_{hr}^i, q_{hp}^i, q_k^i, q_s^i, q_t^i, q_{hs}^i, q_{tr}^i, q_{tp}^i, q_{ta}^i, q_{tb}^i, q_{sy}^i, q_{sr}^i, q_{sp}^i, q_e^i) \quad (19)$$

For each leg, the 11 configuration variables represent the open-chain configuration and are not all independent degrees of freedom. Due to the three closed-loop kinematic constraints present in each leg, these variables correspond to 8 independent operational degrees of freedom. This nomenclature follows the configuration representation provided by the Digit humanoid’s onboard state estimator and previous literature (Narkhede et al. 2022) on the Digit humanoid platform. The state space of the entire free-floating Digit denoted by $(\mathbf{q}', \mathbf{v}')$ is given by:

$$\mathbf{q}' = (\mathbf{p}_o, \mathbf{Q}_o, \mathbf{q}_L, \mathbf{q}_R) \quad (20)$$

$$\mathbf{v}' = (\dot{\mathbf{p}}_o, \boldsymbol{\omega}_o, \dot{\mathbf{q}}_L, \dot{\mathbf{q}}_R) \quad (21)$$

where $\boldsymbol{\omega}_o$ is the angular velocity of the torso in body frame. Let $\mathbf{Q}_b \in SO(3)$ represent the orientation of the object. The complete state space of Digit with the states of the object is given by

$$\mathbf{q} = (\mathbf{q}', \mathbf{x}^b, \mathbf{Q}_b) \in \mathbb{R}^{44} \quad (22)$$

$$\mathbf{v} = (\mathbf{v}', \dot{\mathbf{x}}^b, \boldsymbol{\omega}_b) \in \mathbb{R}^{42} \quad (23)$$

where $\boldsymbol{\omega}_b$ is the angular velocity of the box in the inertial frame.

2.4.2 Constraints

The literature on whole-body control for pHHI typically treats interaction forces as external disturbances within the system dynamics, without explicitly planning for these forces. In contrast, our approach models the coupled dynamics of the robot (Digit) and the object, enabling the robot to actively regulate the interaction forces. This formulation also provides the flexibility to control both the position and orientation of the object.

In our walking model, locomotion alternates between left and right single-support phases, with the double-support phase assumed to have negligible duration. The dynamics of Digit during the single-support phase are expressed as:

$$\mathbf{D}(\mathbf{q})\dot{\mathbf{v}} + \mathbf{C}(\mathbf{q}, \mathbf{v}) = \mathbf{S}_\tau^\top \boldsymbol{\tau} + (\mathbf{J}(\mathbf{q}))^\top \boldsymbol{\lambda} + (\mathbf{J}_L(\mathbf{q}))^\top \boldsymbol{\lambda}_L + (\mathbf{J}_R(\mathbf{q}))^\top \boldsymbol{\lambda}_R \quad (24)$$

where $\mathbf{D}(\mathbf{q})$ is the inertia matrix, $\mathbf{C}(\mathbf{q}, \mathbf{v})$ is the matrix containing velocity-dependent terms and gravitational forces, and $\mathbf{S}_\tau$ is the selection matrix for the actuated joints. $\mathbf{J}_L(\mathbf{q})$ denotes the Jacobian associated with the ball-joint constraint between the left hand and the object, and $\boldsymbol{\lambda}_L = [\lambda_L^x, \lambda_L^y, \lambda_L^z]^\top \in \mathbb{R}^3$ is the corresponding generalized force. An analogous notation is used for the right hand, replacing the subscript ‘L’ with ‘R’. $\mathbf{J}(\mathbf{q})$ denotes the Jacobian of all holonomic constraints not involving the object, and $\boldsymbol{\lambda} \in \mathbb{R}^{16}$ are the associated generalized forces. These constraints include: (a) six closed-loop kinematic constraints arising from the Digit’s mechanical structure, (b) a stance-foot contact constraint, and (c) stiff spring constraints. The closed-loop kinematic constraints are handled in a manner similar to the approach used for the Cassie robot (Reher and Ames 2021). All these

constraints are incorporated as:

$$\mathbf{J}(\mathbf{q})\dot{\mathbf{v}} + \dot{\mathbf{J}}(\mathbf{q}, \mathbf{v})\mathbf{v} = \mathbf{0} \quad (25)$$

Next, we incorporate the dynamics of the object, which is modeled as a single rigid body. Let α_b and ω_b denote the angular acceleration and angular velocity of the object, respectively, expressed in the world frame. The object's moment of inertia matrix in the body frame is denoted by $\mathbf{I}_b$, and its mass by m_b . Let $\mathbf{F}_h$ and $\mathbf{M}_h$ represent the net forces and moments applied by the human at the object's CoM in the world frame. The rotation matrix $\mathbf{R} \in SO(3)$ transforms quantities from the object frame to the world frame, and $\mathbf{F}_g$ denotes the gravitational force. Finally, let $\mathbf{r}_L$ and $\mathbf{r}_R$ be the vectors, expressed in the world frame, from the object's CoM to the left-hand and right-hand contact points, respectively. The dynamics of the object can then be expressed as:

$$\begin{aligned} \mathbf{M}_h - \mathbf{r}_L \times \lambda_L - \mathbf{r}_R \times \lambda_R &= \mathbf{R}\mathbf{I}_b\mathbf{R}^T\alpha_b \\ &\quad + \omega_b \times \mathbf{R}\mathbf{I}_b\mathbf{R}^T\omega_b \end{aligned} \quad (26)$$

$$m_b\ddot{\mathbf{x}}^b = \mathbf{F}_g - \lambda_L - \lambda_R + \mathbf{F}_h \quad (27)$$

Finally, we add 2 ball joint constraints between the two arms and the object:

$$\begin{aligned} \mathbf{J}_i(\mathbf{q})\dot{\mathbf{v}} + \dot{\mathbf{J}}_i(\mathbf{q}, \mathbf{v})\mathbf{v} &= \ddot{\mathbf{x}}^b + \mathbf{r}_i \times \alpha_b + \\ &\quad \omega_b \times (\omega_b \times \mathbf{r}_i) \quad \forall i \in \{L, R\} \end{aligned} \quad (28)$$

This completes the constraints regarding modeling the coupled system. Next, we require the foot contact wrench, $(\lambda^x, \lambda^y, \lambda^z, \lambda^{mx}, \lambda^{my}, \lambda^{mz})$ to lie inside the linearized Contact Wrench Cone (CWC) $\mathcal{K}$ (Caron et al. 2015). We also bound the input torque and interaction forces at the hands as:

$$\mathbf{lb}_\tau \leq \boldsymbol{\tau} \leq \mathbf{ub}_\tau \quad (29)$$

$$\mathbf{lb}_h \leq \lambda_i \leq \mathbf{ub}_h \quad \forall i \in \{L, R\} \quad (30)$$

Here, $\mathbf{lb}_\tau$ and $\mathbf{ub}_\tau$ are the lower and upper bounds on the input torque, $\boldsymbol{\tau}$. Moreover, $\mathbf{lb}_h$ and $\mathbf{ub}_h$ are the lower and upper bounds on the interaction forces between the box and the hand.

2.4.3 Trajectory Tracking Objectives

We now describe the objectives related to tracking the desired trajectories of various quantities

of the Digit humanoid. These quantities include the trajectory of the swing foot, the orientation of the swing foot, the orientation of the torso, and the partial orientation of the object. Let the errors between the desired and actual values of these quantities be stacked into the vector $\mathbf{e} \in \mathbb{R}^{11}$. The corresponding tracking objective, commonly adopted in the literature (Narkhede et al. 2022), is to minimize the cost function Ψ_1 , defined as:

$$\Psi_1 = \|\ddot{\mathbf{e}} + \mathbf{K}_D\dot{\mathbf{e}} + \mathbf{K}_P\mathbf{e}\|^2 \quad (31)$$

where, $\mathbf{K}_D$ and $\mathbf{K}_P$ are appropriately selected gain matrices.

To align the torso orientation with the estimated intended yaw of the object, we define the desired torso orientation as

$$\mathbf{Q}_{\text{des}}^{\text{torso}} = Q(0, 0, \theta_d^b),$$

where $Q(\cdot)$ denotes the function that converts roll-pitch-yaw angles into a quaternion representation.

Using the MPC plan for the next step ($\mathbf{u}_k, \theta_a^s(k)$), we get the desired swing foot trajectory as:

$$\mathbf{p}_{\text{sw}}^{\text{des}}(s) = \begin{bmatrix} \left(\frac{1+\cos \pi s}{2}\right) p_{\text{sw}}^x(k-1) + \left(\frac{1-\cos \pi s}{2}\right) u_k^x \\ \left(\frac{1+\cos \pi s}{2}\right) p_{\text{sw}}^y(k-1) + \left(\frac{1-\cos \pi s}{2}\right) u_k^y \\ P(s, z_{\text{cl}}) \end{bmatrix} \quad (32)$$

Here, s is a phase variable linearly varying from 0 at the start of the step to 1 at the end of the step. P is a fourth-order polynomial in s , designed to ensure that the swing foot starts with zero initial vertical velocity and zero initial vertical position with respect to the stance foot, while attaining a maximum height of z_{cl} during the step. $\mathbf{p}_{\text{sw}}(k) = [p_{\text{sw}}^x(k) \ p_{\text{sw}}^y(k) \ p_{\text{sw}}^z(k)]^T$ is the swing foot position at the start of the k^{th} step. For the orientation of the swing foot, we set the desired orientation $\mathbf{Q}_{\text{des}}^{\text{sw}} = Q(0, 0, \theta_a^s(k))$, where $Q(\cdot)$ converts roll-pitch-yaw angles into quaternion form.

The robot is selectively commanded to control the roll and pitch of the object, maintaining both at a desired value of zero. All orientation errors are expressed in the global frame. The desired quaternion for the object is defined as $Q(0, 0, 0)$; however, the weight associated with the z -axis

rotation is set to zero. This formulation implicitly governs the object's vertical position based on the human's input: when the human raises or lowers the box, a pitch error is detected, prompting the robot to adjust the height accordingly. Rotation about the z -axis is handled separately through a dedicated objective that enforces compliance in yaw.

2.4.4 Other Objectives

We impose objectives on both the humanoid and object center-of-mass (CoM) accelerations to achieve the compliant behavior prescribed by the admittance model. This formulation allows the robot's arms to actively contribute to compliance, complementing the stepping pattern generated by the MPC. Let $\mathbf{a}^c$ and $\mathbf{a}^b$ denote the desired CoM accelerations of the humanoid and the object, respectively. These quantities are obtained from the admittance model depicted in Fig. 2(b) and are computed as:

$$\begin{aligned} m_c \mathbf{a}^c &= \mathbf{R}(\theta^s) \mathbf{K}_h (\mathbf{R}(\theta^s))^\top (\mathbf{x}^b - \mathbf{x}^c - \mathbf{x}^d) \\ &\quad + \mathbf{R}(\theta^s) \mathbf{B}_h (\mathbf{R}(\theta^s))^\top (\dot{\mathbf{x}}^b - \dot{\mathbf{x}}^c) \end{aligned} \quad (33)$$

$$\begin{aligned} m_b \mathbf{a}^b &= \mathbf{F}_h - \mathbf{R}(\theta^s) \mathbf{K}_h (\mathbf{R}(\theta^s))^\top (\mathbf{x}^b - \mathbf{x}^c - \mathbf{x}^d) \\ &\quad - \mathbf{R}(\theta^s) \mathbf{B}_h (\mathbf{R}(\theta^s))^\top (\dot{\mathbf{x}}^b - \dot{\mathbf{x}}^c) \end{aligned} \quad (34)$$

Here, $\mathbf{K}_h = \text{diag}(K_h^x, K_h^y)$ and $\mathbf{B}_h = \text{diag}(B_h^x, B_h^y)$ are different from those in the admittance model, as we hypothesize that low-compliance desired trajectories at the high level and high-compliance desired trajectories at the low level lead to more efficient collaboration. We prove this hypothesis in the results section of this paper. The apparent inertia experienced by the human can be modified by adjusting the masses of the object and the robot in these equations. The tracking objective for the humanoid CoM acceleration is defined as

$$\Psi_2 = \|\ddot{\mathbf{x}}^c - \mathbf{a}^c\|^2, \quad (35)$$

while the corresponding objective for the object CoM acceleration is given by

$$\Psi_3 = \|\ddot{\mathbf{x}}^b - \mathbf{a}^b\|^2. \quad (36)$$

In most frameworks, the desired CoM acceleration for the low-level controller is derived from

the simplified model used for planning (Kumbhar and Artemiadis 2025). In this work, however, we instead use the accelerations from the admittance model as the desired CoM accelerations, thereby promoting *local* compliance. This choice is motivated by the fact that the high-level planner enforces a desired compliance objective, whereas the stiffness modulation only attempts to track the desired distance between the humanoid and the object. Consequently, the desired accelerations generated by the I-LIP model do not exclusively promote compliance. We have similar costs on the yaw acceleration of the object given by:

$$\Psi_4 = \|\ddot{\theta}^b - \alpha^b\|^2. \quad (37)$$

Here, α^b is the desired acceleration given by the admittance model as:

$$\alpha^b = k_p^b(\theta^s - \theta^b) - k_D^b \dot{\theta}^b \quad (38)$$

Here, k_p^b and k_D^b are positive constants for modulating the stiffness in the rotational z direction. We add the small damping term ($k_D^b \dot{\theta}^b$) to stabilize the response. The CoM in the I-LIP model is represented as a point mass and, therefore, cannot capture the effects of angular momentum on Digit. We, therefore, minimize the angular momentum as done by Narkhede et al. (2022). We compute the angular momentum of the digit about its CoM, $\eta_2 = \mathbf{A}_{\text{ang}} \mathbf{v}$. Here, $\mathbf{A}_{\text{ang}}$ is the part of the centroidal momentum matrix. The minimization objective is given by:

$$\Psi_5 = \|\dot{\eta}_2 + \mathbf{K}_{\text{ang}} \eta_2\|^2. \quad (39)$$

$\mathbf{K}_{\text{ang}}$ is appropriately selected gain matrix. Next, we also minimize the input torque as:

$$\Psi_6 = \|\tau\|^2 \quad (40)$$

Finally, we minimize the x and y direction forces at the point of interaction between the box and the arms of the robot.

$$\Psi_7 = \|\lambda_{xy}^L\|^2 + \|\lambda_{xy}^R\|^2 \quad (41)$$

The coupled dynamics defined by (24), (26), and (27) enable direct minimization and bounding of interaction forces via (41) and (30).

2.4.5 Weighted QP

To formulate the optimization problem, we construct a single quadratic cost function by taking a weighted sum of the objectives described above. Together with the system constraints, this cost function defines the QP, expressed as:

$$\begin{aligned} & \underset{\dot{\mathbf{v}}, \ddot{\mathbf{x}}_b, \alpha_b, \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\lambda}^L, \boldsymbol{\lambda}^R}{\text{minimize}} && \sum_{i=1}^7 \psi_i \Psi_i && (42) \\ \text{subject to} && && (24), (25), (26), (27), (28), (29), (30) \\ && && (\lambda^x, \lambda^y, \lambda^z, \lambda^{mx}, \lambda^{my}, \lambda^{mz}) \in \mathcal{K} && (43) \end{aligned}$$

The weighting gains ψ_i , $\forall i \in \{1, 2, 3, 4, 5, 6, 7\}$ are chosen based on the relative priority of the corresponding objectives. The final constraint enforces that the contact wrench at the stance foot lies within the linearized contact wrench cone $\mathcal{K}$.

2.5 Stiffness Modulation

The springs and dampers in the I-LIP model apply forces to the robot's CoM during the footstep planning phase, thereby adjusting the planned footstep pattern based on these interaction forces. We hypothesize that using constant spring and damper parameters will not yield the desired compliant behavior between the object and the robot, due to discrepancies between the forces actually experienced by the CoM and those predicted by the I-LIP model. Kumbhar and Artemiadis (2025) prove this hypothesis for straight line walking. However, their approach employs a global trajectory from the admittance model for stiffness modulation, which does not produce *local* compliance. In this work, we modify the stiffness modulation to achieve convergence to a desired distance between the humanoid and the object rather than to a global trajectory. While this does not fully realize local compliance, the modulation now operates toward convergence to the desired distance (x_x^d) prescribed by the local compliance objective. Additionally, we adopt this strategy for considering turning maneuvers as well.

The modulation of the stiffness of the x -component spring at each iteration of the low-level control loop is given by:

$$\begin{aligned} K_t^x &= K_t^x - k_1^x (x_t^b(t) - x_t^c(t) - x^d) && (44) \\ &- b_1^x (\dot{x}_t^b(t) - \dot{x}_t^c(t)) \end{aligned}$$

$$\mathbf{x}_l^i = (\mathbf{R}(\theta^s))^T \mathbf{x}^i \quad \dot{\mathbf{x}}_l^i = (\mathbf{R}(\theta^s))^T \dot{\mathbf{x}}^i \quad i \in \{b, c\} \quad (45)$$

During hardware experimentation, the initial stiffness was set to $K_0^x = 103.4$ N/m, while the stiffness modulation gains were chosen as $k_1^x = 0.235$ N/m² and $b_1^x = 2.35$ Ns/m². These values were selected through simulation experimentation to achieve stable collaborative behavior. The initial stiffness primarily governs the transient response at the onset of interaction, whereas k_1^x and b_1^x determine the trade off between maintaining the desired humanoid-object separation and humanoid object velocity matching.

The objective of this modulation is not to precisely estimate the forces experienced by the CoM, but rather to adjust them such that the stepping pattern generated by the MPC achieves the desired compliant behavior while promoting convergence of the distance between the CoMs of the robot and the object to the target value x_x^d . This approach is analogous to concepts in adaptive control theory, where the primary goal is to accomplish the task while estimating unknown parameters, with the understanding that these estimates may differ from their true values.

3 Results

We validate our proposed framework through both simulation studies and real-world experiments¹ on the humanoid platform Digit. The evaluation spans multiple collaborative locomotion and manipulation tasks, including in-place walking under varying compliance levels, two-dimensional translations, and turning maneuvers. Combinations of these basic maneuvers can be used to realize more complex behaviors such as navigation in cluttered environments or operation within structured workspaces. Through our framework, these maneuvers are autonomously executed without needing thresholding or state machines.

The control framework is first tested in the MuJoCo physics engine (Todorov et al. 2012) and then deployed on the real robot. To reduce the sim-to-real gap, we add actuator dynamics through the MuJoCo plugin, include joint friction and damping, inject observation noise, introduce

¹ A video of the experimental demonstrations is available at: <https://www.youtube.com/watch?v=VeCh9ycYWWA>

delays in control commands, and vary spring parameters in the humanoid model. Together, these measures significantly improved the fidelity of the simulation environment. As a result, controllers designed and validated in MuJoCo required only minimal tuning when transferred to the physical robot. All results presented in this work are therefore based on successful execution on the real Digit platform, with simulation serving as a critical intermediary step for safe, systematic, and resource-efficient validation. In addition, we developed a standing controller for collaborative lifting, but its details are omitted due to space constraints and the focus of this work. A finite state machine governs the transition between standing and walking modes to conduct all experiments presented in this paper.

This version of Digit lacks wrists to emulate ball joints. To provide wrist functionality and enable interaction force measurement, we designed an aluminum box, housing four ATI Axia130 M125 force-torque sensors: two on the human side, terminating in handles, and two on Digit’s side, connected through a ball-joint and gauntlet design, as shown in Fig. 4. Our work relies on direct force-torque sensor measurements on the human side. When direct measurements are not available, these forces and torques can also be estimated using methodologies reported by [Haddadin et al. \(2017\)](#).

The box weighs 15 kg or 33 pounds. Additionally, we use a camera-based motion tracking system and related software (Optitrack Motive 3.1.1) for obtaining the orientation of the box using reflective markers. Box spatial velocity estimates are computed using a Kalman filter driven by pose measurements from the motion capture system. Similarly, the humanoid torso linear velocity is estimated via a Kalman filter using motion capture position data, while the remaining state estimates are provided by digit’s in-built estimator. The Kalman filter gains are tuned to ensure that the noise amplitude of the resulting velocity signals is comparable to that of the built-in estimator. While the reliance on an external motion capture system restricts these physical validation experiments to a laboratory environment, we performed a simulation-based sensitivity analysis to verify that the MPC and the coupled whole-body framework inherently preserve dynamic balance under feedback degradation. Specifically, we

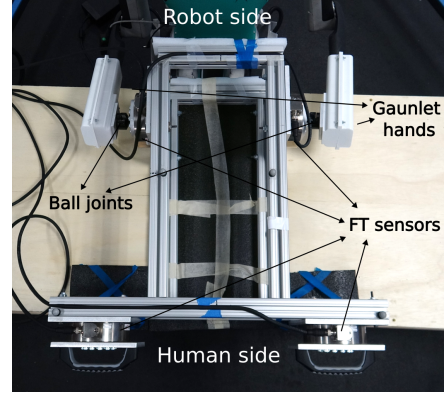


Fig. 4 Physical box used for the collaboration task. The box itself is made of aluminum and weighs around 15 kg. The hands of Digit are connected to the box through a 3D-printed gauntlet and stainless steel ball joint assembly. The interaction ports at the human and robot side also consist of force torque sensors.

injected a 15 ms control input delay and corrupted the motion-capture-derived states with a zero-mean Gaussian noise up to a maximum standard deviation of 0.1 m/s for linear velocities, 0.1 rad/s, and 0.1 rad for orientations. The proposed framework successfully maintained dynamic balance and experienced an approximate 14.8% drop from 0.981 to 0.836 in the task efficiency metric (η). Future work will address the use of onboard camera sensors and the ability to handle the associated noisy signals.

In the following subsections, we present the analysis and main contributions of our framework in the context of co-transportation tasks involving heavy objects. All the analysis done is for the walking controller.

3.1 In-place Walking and Joint Manipulation Evaluation

This task involves a human and the humanoid lifting an object and keeping it in one place. The humanoid is doing in-place walking throughout the experiment. Figure 5 illustrates the progression of the experiment over time through images, with the accompanying plots presenting the analysis and results. Through these experiments, we aim to demonstrate the following aspects of the framework:

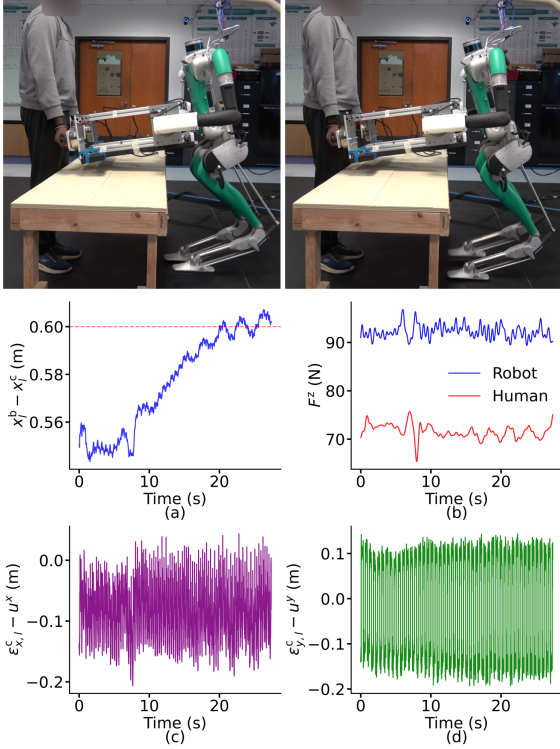


Fig. 5 Results on collaborative in-place walking task. Top: snapshots show the humanoid sharing the load while stepping in place. Bottom: the subscript l in the variables indicates representations in the local frame about the stance foot (X' , Y'). (a) The distance between the CoMs of the humanoid and the object ($x_l^p - x_l^c$) converges to the desired value of 0.6 m. (b) Vertical direction forces applied by the human and the robot (F_h^z , F_r^z) on the object. (c,d) Modified capture point offsets in the locally forward ($\epsilon_{x,l}^c$) and lateral direction ($\epsilon_{y,l}^c$) remain bounded, indicating stability.

Tracking of desired distance from the object

Figure 5(a) shows the distance between the humanoid's CoM and the box converging to the desired 0.6 m over time, demonstrating that the MPC generates footstep patterns that keep the robot close to the box. Stiffness modulation further enhances this behavior by enabling adaptive stepping to achieve convergence to desired distances. The small oscillations observed in the trajectory are primarily caused by the impacts generated during the robot's stance foot switching in in-place walking.

Evaluation of load sharing

Figure 5(b) shows the z-direction forces applied by the humanoid and the human during the task. The humanoid carries more than half of the load, thereby reducing the vertical effort required from the human by more than 50%. The interaction force distribution depends on the relative locations of the contact points with respect to the box center of mass. In our setup, the CoM of the box is located closer to the robot-side interaction ports than to those of the human. Consequently, to maintain static equilibrium while enforcing a zero-pitch objective, the robot naturally supports more than 50% of the load. As a result, the interaction force applied by the human remains consistently lower than that applied by the robot. For collaborative tasks, particularly those difficult for a human to perform alone, such as heavy object transportation or long-duration efforts, effective load sharing reduces human effort, enhances stability, and promotes ergonomics, thereby improving the overall efficiency of the task.

Evaluation of stability through footstep patterns

The concept of a capture point is widely used for controlling legged robots and also serves as a metric for stability and dynamic balance. Boundedness of the capture point about the stance foot indicates stability in the sense of balance. The farther the capture point from the stance foot, the more prone it is to falling. To evaluate stability in the presence of external forces, we use a modification of this capture point to account for the external forces and their effect on the dynamic balance capabilities of the humanoid. The mathematical definitions of capture point emerge from the divergent component of the LIP model. The dynamics of this model in the presence of external forces is given by:

$$m_c \ddot{\mathbf{x}}^c = m_c \omega_o^2 (\mathbf{x}^c - \mathbf{u}) + \mathbf{F}_{ext} \quad (46)$$

Here, $\mathbf{F}_{ext}$ is the equivalent force at the CoM that produces the same moment about the stance foot as the interaction forces at the hands. The above equation can be rearranged as:

$$\ddot{\mathbf{x}}^c = \omega_o^2 \left(\mathbf{x}^c + \frac{\mathbf{F}_{ext}}{m_c \omega_o^2} - \mathbf{u} \right). \quad (47)$$

We define a force and position combined state variable (γ^c) as:

$$\gamma^c = \mathbf{x}^c + \frac{\mathbf{F}_{ext}}{m_c \omega_o^2} \quad (48)$$

Assuming the derivatives of forces ($\mathbf{F}_{ext}$) are zero and substituting γ^c in (47) we get:

$$\ddot{\gamma}^c = \omega_o^2(\gamma^c - \mathbf{u}). \quad (49)$$

These dynamics are similar to the dynamics of the LIP model. Using concepts from the capturability frameworks (Koolen et al. 2012), we define the modified capture point, ξ^c as:

$$\xi^c = \gamma^c + \frac{\dot{\gamma}}{\omega_o} \quad (50)$$

Walking stability and balance can be ensured if the modified capture point is bounded with respect to the stance foot. For all experiments, we compute the *modified capture point offset* denoted by ϵ^c and defined as the relative position of the modified capture point from the stance foot, $\xi^c - \mathbf{u}$ to demonstrate stability in the sense of dynamic balance.

Remark: We do not aim to provide a formal mathematical stability analysis. Instead, stability is evaluated in terms of the ability to generate appropriate footstep patterns to maintain balance and to provide corrective footstep placements in the presence of disturbances or abrupt motions, similar to those presented in Section 3.2.

Figure 5(c,d) shows the plots of this offset in the forward and lateral directions obtained by the transformation $(\mathbf{R}(\theta^s))^T \epsilon^c$. The results demonstrate that the offset exhibits a near-periodic and bounded behavior, which indicates stability/balance. Importantly, it does not diverge, confirming that the MPC generates stepping patterns that maintain stability even in the presence of external interaction forces, thereby preventing falls.

3.2 Evaluation of the Effect of Compliance

There are two primary ways in which compliant behavior can be realized in human-humanoid collaboration: through interactions at the hands and through the stepping pattern. Prior work

by Kumbhar and Artemiadis (2025) has demonstrated the use of an MPC framework to generate stepping patterns that track a desired compliant response. In this work, we extend these capabilities by also leveraging the humanoid’s hands to actively contribute to compliance during co-transportation. Our framework enables the hands and the stepping pattern to execute desired trajectories coming from different levels of compliance given by the stiffness and damping parameters.

We hypothesize that collaboration is most efficient when the stepping pattern follows a relatively stiff trajectory, while the hands exhibit compliant behavior. The compliance at the hands effectively creates a cushioning layer between the humanoid and the box, which not only promotes more stable locomotion but also improves the human’s efficiency in maneuvering the object. On the other hand, prescribing a low-compliance (i.e., stiff) trajectory for the stepping pattern enhances the robot’s reactivity, thereby increasing human efficiency in moving the box. However, when the locomotion speeds are too high or the capture point drifts far from the stance foot, the risk of falling is higher for the case of stiff trajectories.

We prove our hypothesis through a dynamic experiment, where the human as a leader is asked to move the box forward and backward in an approximately periodic fashion as shown in Fig. 6(top). This ensures that the position and velocity of the box change frequently for the compliance properties to show effect. We run four different experiments for all high-compliance and low-compliance combinations of the desired trajectories in the high-level planner and the control

Table 1 Average efficiency ($\bar{\eta}$) for the four cases during the periodic motion task.

Case	Parameters(N/m & Ns/m)	$\bar{\eta}$
1	$K_a^x = 500, B_a^x = 40$ (Low)	0.818
	$K_h^x = 500, B_h^x = 40$ (Low)	
2	$K_a^x = 25, B_a^x = 10$ (High)	0.677
	$K_h^x = 25, B_h^x = 10$ (High)	
3	$K_a^x = 500, B_a^x = 40$ (Low)	0.814
	$K_h^x = 25, B_h^x = 10$ (High)	
4	$K_a^x = 25, B_a^x = 10$ (High)	0.524
	$K_h^x = 500, B_h^x = 40$ (Low)	

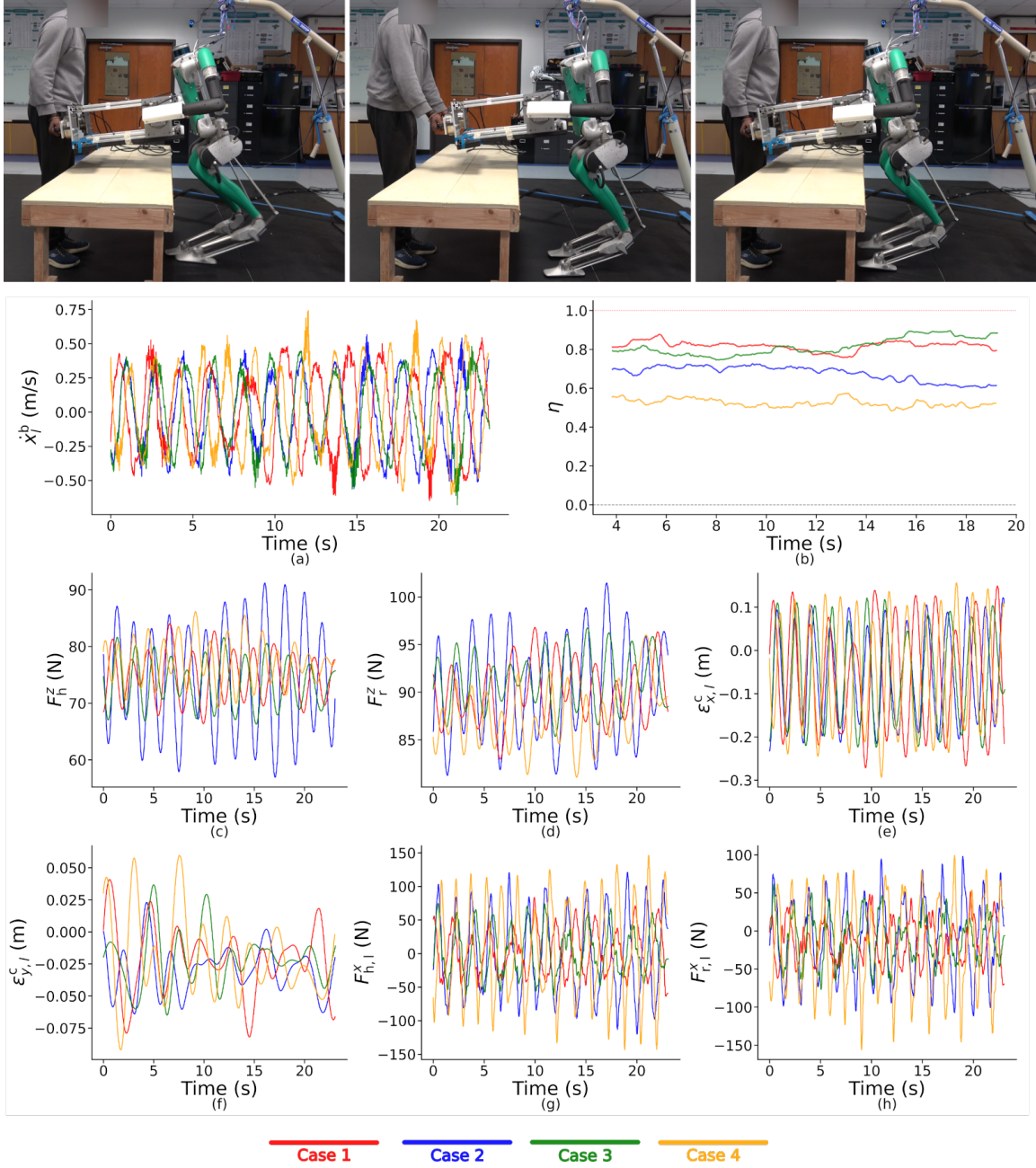


Fig. 6 Top: Experiment illustrating the periodic motion task in which the human moves the box forward and backward in a periodic fashion, as shown in the task progression snapshots. Bottom: The subscript l indicates representations in the local frame about the stance foot (X' , Y'). (a) Box velocity ($\dot{x}_l^b$) in the forward direction for all four cases showing the periodic motion. (b) Efficiency (η) of all cases, showing high efficiencies in cases involving low-compliance objectives for the stepping pattern. (c,d) Human and humanoid's vertical force (F_h^z and F_r^z) on the box, showing more than 50% contribution from the robot. (e,f) Modified capture point offsets in the forward direction ($\epsilon_{x,l}^c$) and the lateral direction ($\epsilon_{y,l}^c$). It remains bounded, indicating that the stepping pattern is able to capture the robot and prevent falls. (g,h) Forces applied by the human (F_h^x) and the humanoid (F_r^x) in the forward direction.

objective for hands in the low-level controller. Table 2 lists the four cases with their respective

parameters. For the following sections, parameters ($\mathbf{K}_i = 25N/m, \mathbf{B}_i = 10Ns/m \forall i \in$

$\{h, a\}$) correspond to high compliance and $(\mathbf{K}_i = 500N/m, \mathbf{B}_i = 40Ns/m \forall i \in \{h, a\})$ correspond to low compliance. Here, instead of keeping the damping parameter constant, we scale it to ensure approximately critically damped behavior. Figure 6(a) shows the locally forward box velocity for all four combinations. All four plots have almost the same amplitude and frequency of oscillation, showing a similar intended motion profile of the human in all four cases of compliance.

To compare the four cases of compliance, we introduce a collaboration efficiency metric. This metric is defined as the ratio between the net effort applied to the box to accomplish the task and the total effort exerted by both the human and the humanoid on the box. Effort is quantified as the time integral of absolute power, capturing the total work performed irrespective of whether it is assistive or resistive. The net effort E_n is given by:

$$E_n = \int_0^w |(\mathbf{F}_h + \mathbf{F}_r) \cdot \dot{\mathbf{x}}^b| dx \quad (51)$$

Here, $\mathbf{F}_h$ and $\mathbf{F}_r$ are the net forces applied by the human and the robot on the box, while w is the time window of the integral, which allows us to quantify how the effort changes over time. The total effort (E_s) is defined as the sum of the individual efforts of the human and the humanoid and it is given by:

$$E_s = \int_0^w (|\mathbf{F}_h \cdot \dot{\mathbf{x}}^b| + |\mathbf{F}_r \cdot \dot{\mathbf{x}}^b|) dx \quad (52)$$

Let η be the efficiency of the collaboration defined as the ratio of net effort to the total effort:

$$\eta = \frac{E_n}{E_s} \quad (53)$$

The proposed metric quantifies the efficiency with which net work is transferred to the object during the collaborative task. A lower efficiency indicates that a significant portion of the individual efforts are either dissipated or cancel each other out, i.e., the agents are working in opposition. This reduces the net work performed on the object. This metric reflects not only task-level performance but also serves as a proxy for the degree of cooperation between the human and the humanoid. Figure 6(b) shows the efficiency calculated over time for the four cases, using a sliding

window (w) of 7.67 s and a stride of 15.34 ms to smooth out high frequencies. As can be seen, the efficiency for Cases 1 and 3, corresponding to a low compliance objective for the stepping pattern, is higher than that of the other cases. We define the mean efficiency, $\bar{\eta}$, as the average value of the efficiency computed using (53) across all sliding windows over the duration of the experiment. Table 2 shows the mean efficiencies ($\bar{\eta}$) for the four cases. Cases 1 and 3 have the highest and comparable efficiencies, supporting our hypothesis that low-compliance footstep objectives play a critical role in enhancing efficiency during collaborative tasks.

In all cases, the humanoid is lifting more than half of the load of the box, as shown in Fig. 6(c,d) by the vertical forces applied by the human and the humanoid. Again, higher load sharing contribution of the humanoid is a consequence of the robot's zero box-pitch objective. Significantly reduced human load enables the co-transportation of heavy objects and the execution of long-duration tasks. Figure 6(e,f) shows the modified capture point offsets for the local X' or forward axis, $\epsilon_{x,l}^c$, and Y' or lateral axis, $\epsilon_{y,l}^c$. As can be seen, the offset stays bounded in both directions, indicating that the robot is not falling and is able to generate a stepping pattern to maintain balance. Additionally, for the lateral direction, it is similar or less in amplitude compared to the in-place walking task, indicating less correlation between the motion in the forward and lateral directions. Perturbations in the forward direction do not affect stability in the lateral direction. Figures 6(g,h) show the forward-direction forces exerted on the box by the human and the humanoid. The framework demonstrates robustness by maintaining balance even under external forces exceeding 100N, with peak values approaching 150N.

3.3 Straight Line Walking Evaluation

Using this framework, we were able to achieve varying forward speeds for the box carried by the human-humanoid pair, with a maximum of approximately 0.7 m/s. Figure 7 shows the different speeds achieved by the box during the collaboration. Under those conditions, we analyzed

the efficiency and robot behavior under high-compliance ($K_h^x = 25.0\text{N/m}$, $B_h^x = 10.0\text{Ns/m}$) and low-compliance ($K_h^x = 500.0\text{N/m}$, $B_h^x = 40.0\text{Ns/m}$) objectives for the hands, while keeping the footstep pattern objective fixed, to highlight the contribution of the hands to efficient collaboration. Figure 8(a) shows the locally forward box velocity for both cases. Both plots have almost the same trajectory, showing a similar intended motion profile of the box in both cases of compliance.

For effort calculations in the efficiency metric, we consider only forward-direction forces and velocities, reflecting the predominantly straight walking nature of the task. Specifically, the dot product computed in (51) and (52) converts to a scalar product between the local x or forward axis force measurements (F_h^x and F_r^x) and the box velocity ($\dot{x}^b$).

Figure 8(b) shows the efficiency calculated over time for both cases. The efficiency for the case of high compliance is consistently higher than that of low compliance. This implies that high hand compliance enables more efficient collaboration, with the dyad working in synchrony to accomplish the task. Figure 8(c,d) shows the forces applied by the human and the humanoid in the vertical direction on the box. Similar to previous experiments, the humanoid contributes to more than 50% of the load, significantly reducing human effort. Figure 8(e,f) shows the evolution of the modified capture point offset, which remains bounded in both the forward and the lateral directions without diverging. Figure 8(g,h) shows the forces applied by the human and the humanoid on the box in the local forward direction. These forces remain minimal,

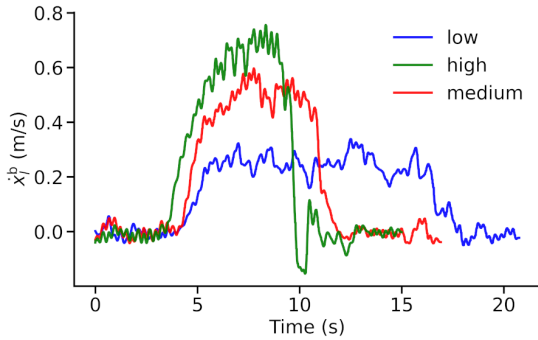


Fig. 7 Different speeds of the box in the forward direction ($\dot{x}_f^b$), low, medium and high, during the straight line walking task.

indicating that the box is easily manipulated by the human.

3.4 Complex Maneuvering Evaluation

The proposed framework enables the humanoid to execute complex maneuvers, including lateral walking and turning. In the pure translation task shown in Fig. 9 (left), the robot sequentially moves forward, laterally left, laterally right, and finally backward to its original position, passively following the human's lead. Figure 9 (right) illustrates a semicircular task that combines translation and yaw rotation, where the human moves the box along a semicircular path and the robot follows passively to complete the motion. It is important to note that in all tasks, the humanoid has no prior knowledge of the human's intended trajectory; it simply responds by following the motion of the object.

Additionally, we validate our compliance hypothesis in simulation for complex, multi-directional co-transportation tasks. Equivalent real-world validation is challenging because human intent continuously evolves to maintain balance, making it difficult to reproduce identical payload trajectories across different compliance settings for a fair efficiency comparison. Furthermore, task efficiency (η) depends jointly on stability and compliance. To isolate the effects of compliance, we designed an experiment where the human is modeled as a damper and its linear and angular velocity commands are modulated by ± 0.3 m/s every 2 s over a 16 s horizon. Additionally, we set the desired distance between the CoMs of the humanoid and the object (x^d) to 0.5 m. This velocity change provides sufficient dynamic excitation to highlight compliance differences without inducing aggressive instabilities that would dominate the metric. As a result, the baseline balance stability across all tested compliance configurations (shown in Table 2) remains tightly matched, as evidenced by the overlapping modified capture point distributions in Fig. 10.

As summarized in Table 2, the 'stiff legs, soft hands' paradigm (Case 3) yields the highest mean collaborative efficiency of 0.753. Low leg compliance ensures high reactivity to box motions, while high arm compliance buffers the humanoid from destabilizing interaction forces. Notably, when the

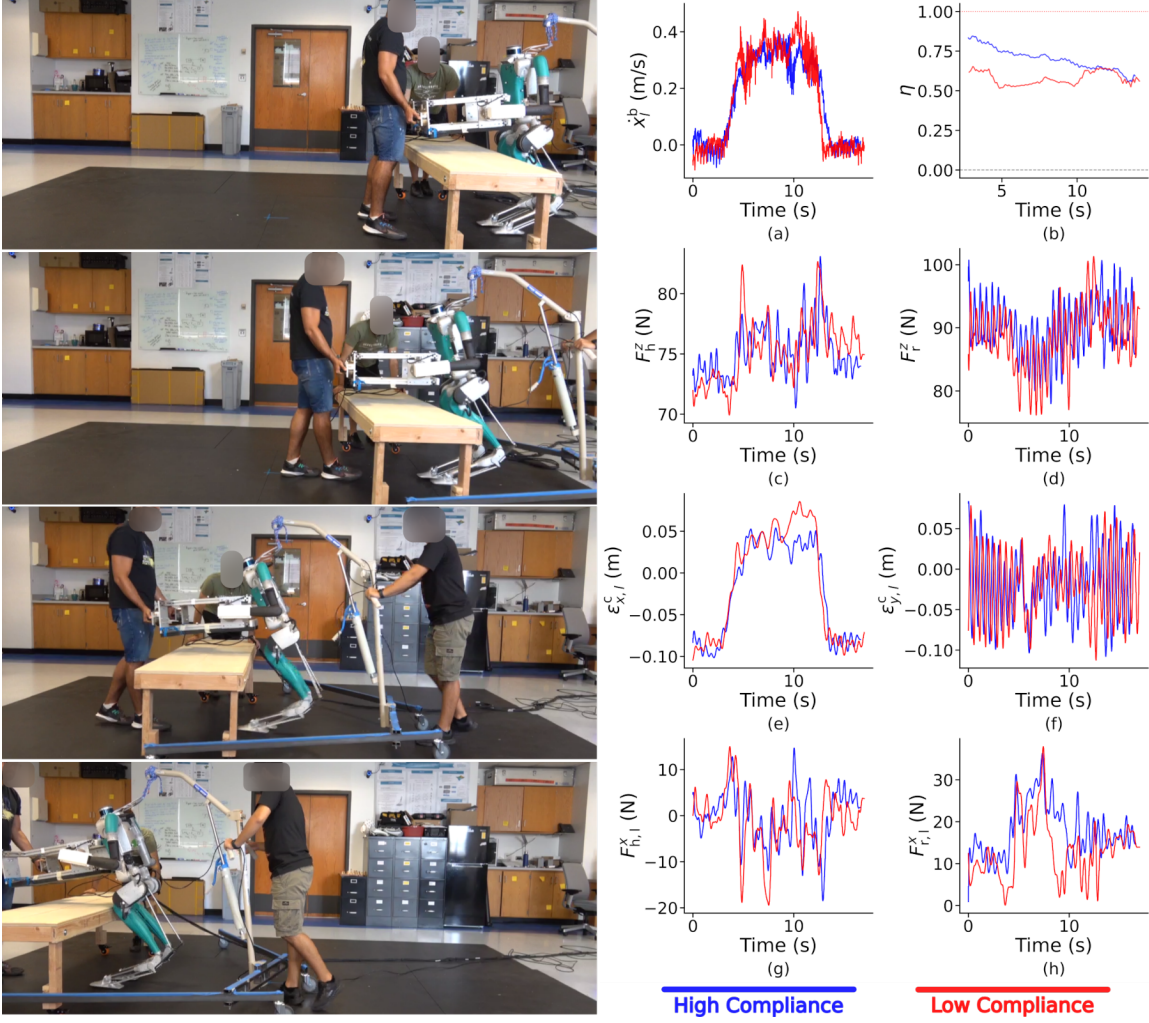


Fig. 8 Experiment illustrating the straight walking task. Left: The human starts moving the box with an intended speed, and the humanoid follows it as shown in the progression through images. Right: We analyze two cases, high and low compliance in hands, for around the same velocity task. The subscript l indicates representations in the local frame about the stance foot (X' , Y'). (a) Box velocity ($\dot{x}_l^b$) in the forward direction for both cases. (b) Efficiency (η) of the task shows consistently higher values for high compliance. (c,d) Vertical forces applied by the human and the humanoid on the object (F_h^z and F_r^z), showing the load sharing. (e,f) Modified capture point offsets in the forward direction ($\epsilon_{x,l}^c$) and the lateral direction ($\epsilon_{y,l}^c$). (g,h) Forces exerted by the human and the humanoid on the box in the forward direction ($F_{h,l}^x$ and $F_{r,l}^x$), which remain minimal.

robot enters highly transient, near-unstable states, low arm compliance can temporarily support stability and efficiency, suggesting a compelling direction for future work in online, stability-driven compliance parameter adaptation.

3.5 Comparative Evaluation of Local and Global Compliance Frameworks

Previous works by [Agravante et al. \(2019, 2016\)](#) do not incorporate compliance between the humanoid and the payload. [Kumbhar and Artemiadis \(2025\)](#) addressed this by incorporating admittance model states directly into the MPC footstep planner.

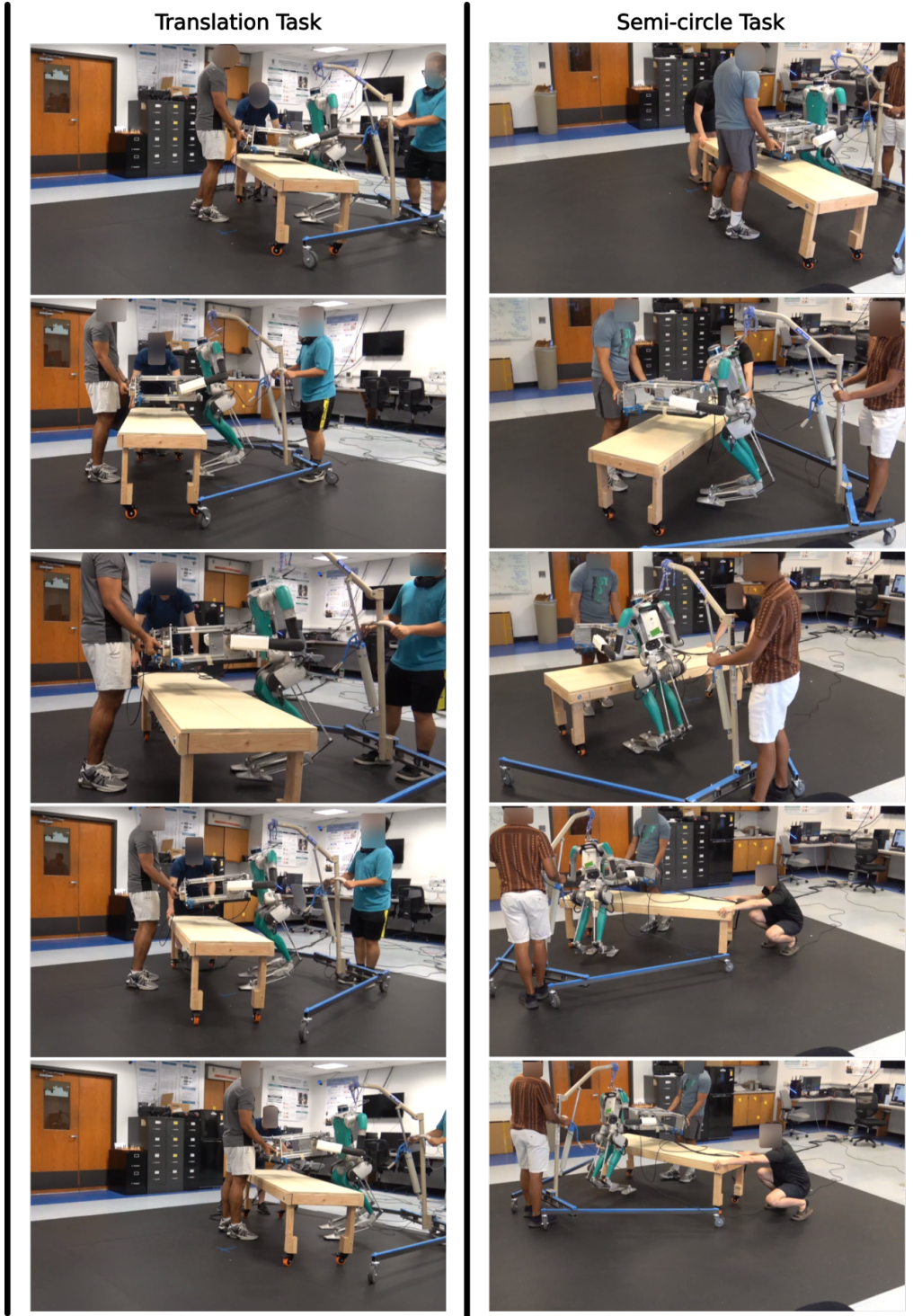


Fig. 9 Progressions of the translational (left) and semi-circle task (right) over time. The humanoid is able to follow complex trajectories for co-transportation tasks.

However, the resulting “global compliance” framework did not include whole-body arm task objectives to explicitly promote local compliance and

enhance the dynamic balance of the humanoid.

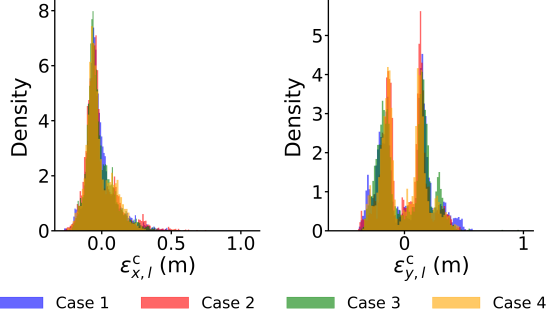


Fig. 10 Distributions of the modified capture point offsets in the forward direction ($\epsilon_{x,l}^c$) and the lateral direction ($\epsilon_{y,l}^c$) for all four cases of compliance settings.

Table 2 Average efficiency ($\bar{\eta}$) for the four cases during the complex motion task.

Case	Parameters(N/m & Ns/m)	$\bar{\eta}$
1	$K_a^x = 500, B_a^x = 40$ (Low)	0.718
	$K_h^x = 500, B_h^x = 40$ (Low)	
2	$K_a^x = 25, B_a^x = 10$ (High)	0.701
	$K_h^x = 25, B_h^x = 10$ (High)	
3	$K_a^x = 500, B_a^x = 40$ (Low)	0.753
	$K_h^x = 25, B_h^x = 10$ (High)	
4	$K_a^x = 25, B_a^x = 10$ (High)	0.666
	$K_h^x = 500, B_h^x = 40$ (Low)	

To evaluate the benefits of the proposed architecture, we present a comparative simulation study in MuJoCo against the “global compliance” framework proposed by [Kumbhar and Artemiadis \(2025\)](#), which serves as a representative baseline.

In this comparative simulation, the human partner is modeled as a high-gain damping controller tasked with maintaining zero payload velocity. During a stable in-place walking sequence, an abrupt force impulse is introduced through the human controller to emulate an unintended external perturbation, after which the human partner immediately resumes the damping behavior. This impulse introduces a velocity spike as shown in Fig. 11(a). We analyze the humanoid’s response during the recovery phase following the perturbation, focusing on recovery efficiency and dynamic stability. The proposed

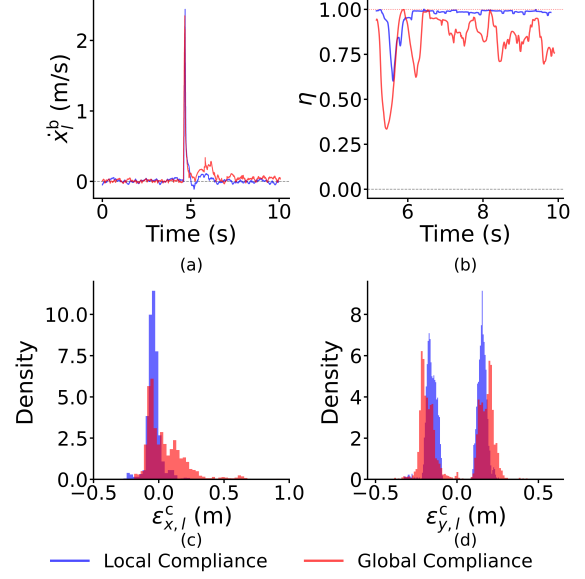


Fig. 11 (a) Box velocity ($\dot{x}_l^b$) in the forward direction for global and local compliance frameworks. (b) Task efficiency (η) during the post-perturbation recovery phase. (c,d) Distributions of the modified capture point offsets in the forward direction ($\epsilon_{x,l}^c$) and the lateral direction ($\epsilon_{y,l}^c$).

framework achieves higher recovery-phase efficiency compared to the baseline, as shown in Fig. 11(b). Furthermore, Fig. 11(c,d) illustrates the spatial distribution of the modified capture point for both frameworks. The baseline framework exhibits a wider spread in the modified capture point distribution after the impulse, indicating reduced dynamic stability compared to the tighter distribution maintained by the proposed method. In the baseline framework, the payload was regulated primarily through a height objective, resulting in excessive compliance during planar motions. Consequently, the relative distance between the humanoid and the payload increased significantly, leading to larger destabilizing effects following the perturbation. In contrast, the improved recovery performance of the proposed framework can be attributed to the integration of local compliance objectives for the arms within the whole-body controller, which enables better regulation of interaction forces while maintaining humanoid balance.

4 Conclusion

We presented an MPC-QP control framework with stiffness modulation to enable

physical human–humanoid collaboration in co-transportation tasks. The framework incorporates a high-level planning layer based on an admittance and an I-LIP driven MPC to generate compliant motion plans, which are executed on the humanoid Digit through a low-level QP that accounts for both robot and object dynamics. Stiffness modulation was introduced to adaptively regulate coupling between the robot and the object, ensuring convergence to the desired relative configuration while maintaining flexibility in interaction.

The proposed approach was validated through real-world experiments on Digit, where we introduced an efficiency metric to jointly assess task performance and inter-agent coordination. This metric was further evaluated using experimental data, demonstrating that the proposed “stiff legs, soft hands” compliance strategy, characterized by low-compliance stepping objectives and high-compliance arm objectives, achieves the highest collaborative efficiency. The experimental results further demonstrate that the proposed framework enables a wide range of co-transportation behaviors, including forward motion, lateral translation, and turning maneuvers. It allows the humanoid to passively adapt to human intent without requiring prior knowledge of the desired trajectory.

This work makes three main contributions: (i) a hierarchical control framework that integrates stiffness modulation into MPC-QP layers for human–humanoid co-transportation, (ii) an efficiency metric for quantitatively evaluating collaboration, and (iii) an experimental demonstration on a full-scale humanoid robot validating the effectiveness of compliance in enhancing dyadic transport and showcasing the robot’s capability to perform diverse and complex co-transportation tasks.

While the proposed framework demonstrates effective co-transportation with planar payload motions, its current formulation maintains the payload roll and pitch angles at zero to ensure a stable and level transport configuration. Future work will extend the framework toward general spatial manipulation tasks by incorporating human intent estimation and whole-body motion adaptation to enable coordinated multi-axis payload rotations. Additionally, the current approach relies on fixed compliance parameters selected for the considered tasks. Future extensions will

investigate online compliance adaptation strategies that adjust these parameters based on real-time stability metrics and task states, enabling improved performance across a broader range of dynamic interaction scenarios. Addressing these limitations will further enhance the versatility and applicability of humanoid robots as collaborative partners in dynamic real-world scenarios.

References

- Agravante, D.J., Cherubini, A., Bussy, A., Gergondet, P., Kheddar, A.: Collaborative human-humanoid carrying using vision and haptic sensing. 2014 *IEEE International Conference on Robotics and Automation (ICRA)*, 607–612 (2014)
- Agravante, D.J., Cherubini, A., Sherikov, A., Wieber, P.-B., Kheddar, A.: Human-humanoid collaborative carrying. *IEEE Transactions on Robotics* **35**(4), 833–846 (2019)
- Agravante, D.J., Sherikov, A., Wieber, P.-B., Cherubini, A., Kheddar, A.: Walking pattern generators designed for physical collaboration. In: 2016 *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1573–1578 (2016). <https://doi.org/10.1109/ICRA.2016.7487296>
- Ajoudani, A., Zanchettin, A.M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., Khatib, O.: Progress and prospects of the human–robot collaboration. *Autonomous robots* **42**(5), 957–975 (2018)
- Bruzzzone, L., Quaglia, G.: Review article: Locomotion systems for ground mobile robots in unstructured environments. *Mechanical Sciences* **3**, 49–62 (2012) <https://doi.org/10.5194/ms-3-49-2012>
- Caron, S., Kheddar, A., Tempier, O.: Stair climbing stabilization of the HRP-4 humanoid robot using whole-body admittance control. In: 2019 *International conference on robotics and automation (ICRA)*, pp. 277–283 (2019). IEEE
- Caron, S., Pham, Q.C., Nakamura, Y.: Stability of surface contacts for humanoid robots: Closed-form formulae of the contact wrench cone for rectangular support areas. *Proceedings*

- *IEEE International Conference on Robotics and Automation* **2015** (2015) <https://doi.org/10.1109/ICRA.2015.7139910>
- Chand, P., Veer, S., Poulakakis, I.: Interactive dynamic walking: Learning gait switching policies with generalization guarantees. *IEEE Robotics and Automation Letters* **7**(2), 4149–4156 (2022)
- Ding, Y., Khazoom, C., Chignoli, M., Kim, S.: Orientation-aware model predictive control with footstep adaptation for dynamic humanoid walking. In: *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pp. 299–305 (2022). <https://doi.org/10.1109/Humanoids53995.2022.10000244>
- DelPreto, J., Rus, D.: Sharing the load: Human-robot team lifting using muscle activity. In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 7906–7912 (2019). <https://doi.org/10.1109/ICRA.2019.8794414>
- Edsinger, A., Kemp, C.C.: Human-robot interaction for cooperative manipulation: Handing objects to one another. In: *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1167–1172 (2007). <https://doi.org/10.1109/ROMAN.2007.4415256>
- Haddadin, S., De Luca, A., Albu-Schäffer, A.: Robot collisions: A survey on detection, isolation, and identification. *IEEE Transactions on Robotics* **33**(6), 1292–1312 (2017)
- Huber, M., Rickert, M., Knoll, A., Brandt, T., Glasauer, S.: Human-robot interaction in handing-over tasks. In: *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 107–112 (2008). <https://doi.org/10.1109/ROMAN.2008.4600651>
- Jarrassé, N., Charalambous, T., Burdet, E.: A framework to describe, analyze and generate interactive motor behaviors. *PLOS ONE* **7**(11), 1–13 (2012) <https://doi.org/10.1371/journal.pone.0049945>
- Kumbhar, S.S., Artemiadis, P.: MPC-QP-based control framework for compliant behavior of humanoid robots in physical collaboration with humans. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14714–14720 (2025). <https://doi.org/10.1109/ICRA55743.2025.11128037>
- Koolen, T., Boer, T., Rebula, J., Goswami, A., Pratt, J.: Capturability-based analysis and control of legged locomotion, part 1: Theory and application to three simple gait models. *The International Journal of Robotics Research* **31**(9), 1094–1113 (2012) <https://doi.org/10.1177/0278364912452673> <https://doi.org/10.1177/0278364912452673>
- Kosuge, K., Kakuya, H., Hirata, Y.: Control algorithm of dual arms mobile robot for cooperative works with human. In: *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace* (Cat.No.01CH37236), vol. 5, pp. 3223–32285 (2001). <https://doi.org/10.1109/ICSMC.2001.972015>
- Kajita, S., Kanehiro, F., Kaneko, K., Yokoi, K., Hirukawa, H.: The 3D linear inverted pendulum mode: A simple modeling for a biped walking pattern generation. In: *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180), vol. 1, pp. 239–2461 (2001). <https://doi.org/10.1109/IROS.2001.973365>
- Kumbhar, S.S., Kulkarni, A.M., Poulakakis, I.: Finite-step capturability and recursive feasibility for bipedal walking in constrained regions. In: *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 397–403 (2025). <https://doi.org/10.1109/ICRA55743.2025.11128831>
- Katayama, S., Murooka, M., Tazaki, Y.: Model predictive control of legged and humanoid robots: Models and algorithms. *Advanced Robotics* **37**(5), 298–315 (2023)
- Kosuge, K., Sato, M., Kazamura, N.: Mobile robot helper. In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on*

- Robotics and Automation*. Symposia Proceedings (Cat. No.00CH37065), vol. 1, pp. 583–5881 (2000). <https://doi.org/10.1109/ROBOT.2000.844116>
- Lefèvre, H., Chaki, T., Kawakami, T., Tanguy, A., Yoshiike, T., Kheddar, A.: Humanoid-human sit-to-stand-to-sit assistance. *IEEE Robotics and Automation Letters* (2024)
- Lasota, P.A., Fong, T., Shah, J.A., *et al.*: A survey of methods for safe human-robot interaction. *Foundations and Trends® in Robotics* **5**(4), 261–349 (2017)
- Li, W., Li, P., Jin, L., Xu, R., Guo, J., Wang, J.: Compliant-control-based assisted walking with mobile manipulator. *Biomimetics* **9**(2), 104 (2024)
- Mujica, M., Crespo, M., Benoussaad, M., Junco, S., Fourquet, J.-Y.: Robust variable admittance control for human–robot co-manipulation of objects with unknown load. *Robotics and Computer-Integrated Manufacturing* **79**, 102408 (2023) <https://doi.org/10.1016/j.rcim.2022.102408>
- Motahar, M.S., Veer, S., Huang, J., Poulakakis, I.: Integrating dynamic walking and arm impedance control for cooperative transportation. In: 2015 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1004–1010 (2015). IEEE
- Motahar, M.S., Veer, S., Poulakakis, I.: Steering a 3D limit-cycle walker for collaboration with a leader. In: 2017 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5251–5256 (2017). IEEE
- Neumann, C.: Humanoid robots offer both disruption and promise. Here’s why. *World Economic Forum*. Accessed: 2025-08-09 (2025). <https://www.weforum.org/stories/2025/06/humanoid-robots-offer-disruption-and-promise/>
- Narkhede, K.S., Kulkarni, A.M., Thanki, D.A., Poulakakis, I.: A sequential MPC approach to reactive planning for bipedal robots using safe corridors in highly cluttered environments. *IEEE Robotics and Automation Letters* **7**(4), 11831–11838 (2022) <https://doi.org/10.1109/LRA.2022.3204367>
- Nguyen, Q., Sreenath, K.: Safety-critical control for dynamical bipedal walking with precise foot-step placement. *IFAC-PapersOnLine* **48**(27), 147–154 (2015)
- Narkhede, K.S., Thanki, D.A., Kulkarni, A.M., Poulakakis, I.: Overtaking moving obstacles with Digit: Path following for bipedal robots via model predictive contouring control. In: 2023 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8590–8596 (2023). IEEE
- Reher, J., Ames, A.D.: Inverse dynamics control of compliant hybrid zero dynamic walking. In: 2021 *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2040–2047 (2021). <https://doi.org/10.1109/ICRA48506.2021.9560906>
- Rozo, L., Calinon, S., Caldwell, D.G.: Learning force and position constraints in human-robot cooperative transportation. In: The 23rd *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 619–624 (2014). <https://doi.org/10.1109/ROMAN.2014.6926321>
- Radosavovic, I., Kamat, S., Darrell, T., Malik, J.: Learning humanoid locomotion over challenging terrain. *arXiv preprint arXiv:2410.03654* (2024)
- Rapetti, L., Sartore, C., Elobaid, M., Tirupachuri, Y., Draicchio, F., Kawakami, T., Yoshiike, T., Pucci, D.: A control approach for human-robot ergonomic payload lifting. In: 2023 *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7504–7510 (2023). IEEE
- Rapetti, L., Tirupachuri, Y., Ranavolo, A., Kawakami, T., Yoshiike, T., Pucci, D.: Shared control of robot-robot collaborative lifting with agent postural and force ergonomic optimization. In: 2021 *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9840–9847 (2021). IEEE
- Suda, R., Kosuge, K.: Handling of object by

- mobile robot helper in cooperation with a human using visual information and force information. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1102–11072 (2002). <https://doi.org/10.1109/IRDS.2002.1043878>
- Todorov, E., Erez, T., Tassa, Y.: Mujoco: A physics engine for model-based control. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033 (2012). <https://doi.org/10.1109/IROS.2012.6386109>
- Veer, S., Motahar, M.S., Poulakakis, I.: Adaptation of limit-cycle walkers for collaborative tasks: A supervisory switching control approach. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5840–5845 (2017). IEEE
- Vianello, L., Penco, L., Gomes, W., You, Y., Anzalone, S.M., Maurice, P., Thomas, V., Ivaldi, S.: Human-humanoid interaction and cooperation: A review. *Current Robotics Reports* **2**(4), 441–454 (2021)
- Wensing, P.M., Posa, M., Hu, Y., Escande, A., Mansard, N., Prete, A.D.: Optimization-based control for dynamic legged robots. *IEEE Transactions on Robotics* **40**, 43–63 (2024) <https://doi.org/10.1109/TRO.2023.3324580>
- Xing, H., Torabi, A., Ding, L., Gao, H., Li, W., Mushahwar, V.K., Tavakoli, M.: Human-robot collaboration for heavy object manipulation: Kinesthetic teaching of the role of wheeled mobile manipulator. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2962–2969 (2021). <https://doi.org/10.1109/IROS51168.2021.9635910>
- Yokoyama, K., Handa, H., Isozumi, T., Fukase, Y., Kaneko, K., Kanehiro, F., Kawai, Y., Tomita, F., Hirukawa, H.: Cooperative works by a human and a humanoid robot. In: *2003 IEEE International Conference on Robotics and Automation* (Cat. No.03CH37422), vol. 3, pp. 2985–29913 (2003). <https://doi.org/10.1109/ROBOT.2003.1242049>