Rhythm: Learning Interactive Whole-Body Control for Dual Humanoids

Hongjin Chen[1,2,*] Wei Zhang[2,*] Pengfei Li[2,3,†] Shihao Ma[1,2] Ke Ma[1,2] Yujie Jin[2] Zijun Xu[1,5]
Xiaohui Wang[1,2] Yupeng Zheng[6] Zining Wang[2] Jieru Zhao[4] Yilun Chen[2] Wenchao Ding[1,2,†]

[1]Fudan University [2]TARS [3]Tsinghua University [4]Shanghai Jiao Tong University
[5]Shanghai Innovation Institute [6]Institute of Automation, Chinese Academy of Sciences

[*]Equal contribution [†]Corresponding Authors

Page: https://hoshi-no-ai.github.io/Rhythm/

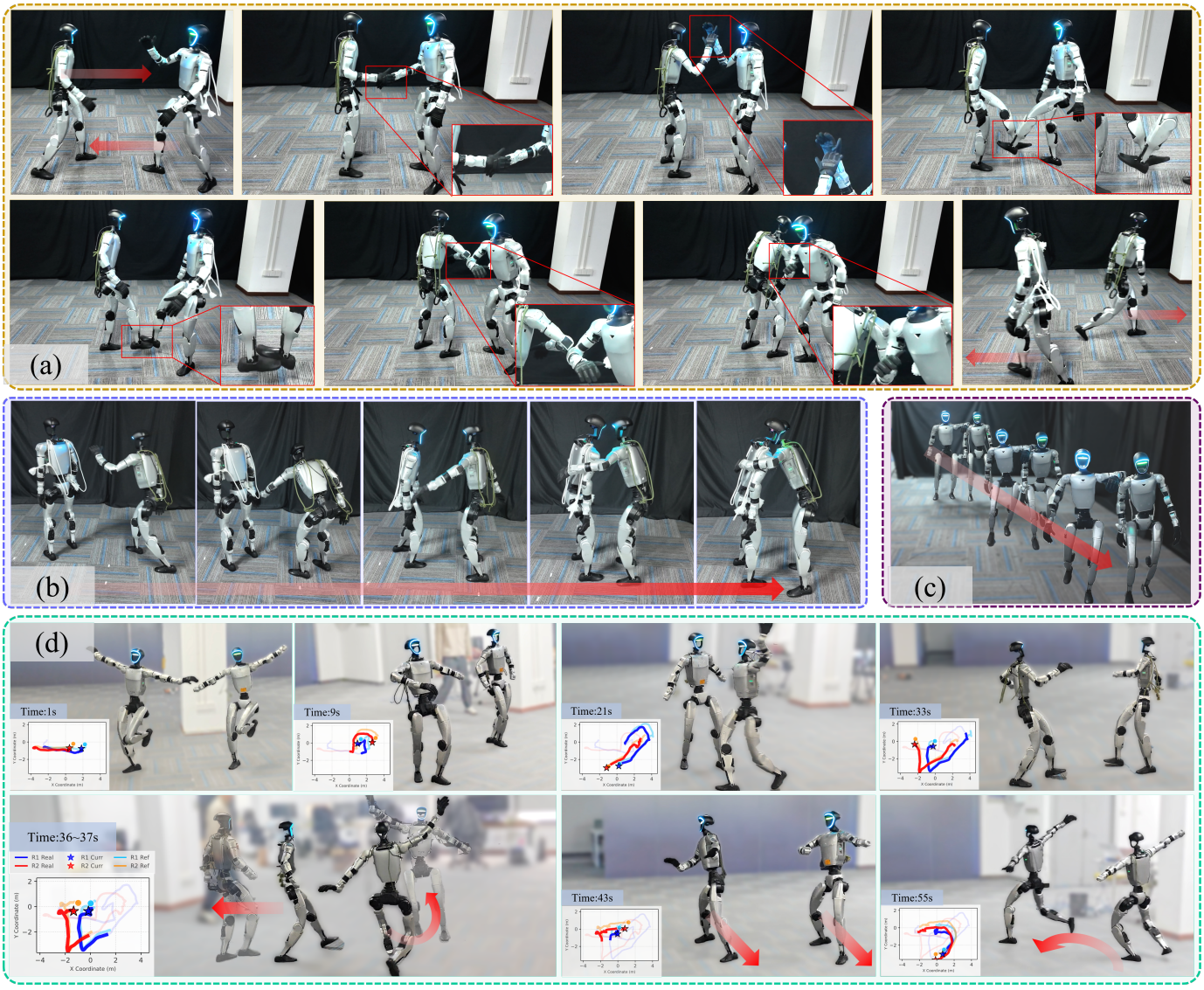

Fig. 1: The proposed framework, **Rhythm**, facilitates a spectrum of humanoid–humanoid interactions. **(a–c) Contact-Rich Interaction:** The method handles interactions ranging from light contact (Greeting) to intensive contact (Hug, Shoulder-to-Shoulder), maintaining fine-grained contact geometry without penetration (shown in the zoomed-in views). **(d) Coordinated Interaction:** The humanoids perform synchronized long-horizon dance (*La La Land*), with trajectories showing consistent spatiotemporal alignment and stable relative positioning over time.

***Abstract*—Realizing interactive whole-body control for multi-humanoid systems is critical for unlocking complex collaborative capabilities in shared environments. Although recent advancements have significantly enhanced the agility of individual robots, bridging the gap to physically coupled multi-humanoid interaction remains challenging, primarily due to**

severe kinematic mismatches and complex contact dynamics. To address this, we introduce Rhythm, the first unified framework enabling real-world deployment of dual-humanoid systems for complex, physically plausible interactions. Our framework integrates three core components: (1) an Interaction-Aware Motion Retargeting (IAMR) module that generates feasible humanoid interaction references from human data; (2) an Interaction-Guided Reinforcement Learning (IGRL) policy that masters coupled dynamics via graph-based rewards; and (3) a real-world deployment system that enables robust transfer of dual-humanoid interaction. Extensive experiments on physical Unitree G1 robots demonstrate that our framework achieves robust interactive whole-body control, successfully transferring diverse behaviors such as hugging and dancing from simulation to reality.

I. Introduction

Humanoid robotics has witnessed rapid evolution, achieving remarkable milestones in single-agent capabilities. Recent research has established a strong foundation in dynamic locomotion [12, 19, 23, 40, 48] and general whole-body control [8, 17, 22, 26, 36, 55], significantly enhancing the agility and robustness of individual robots. However, the broader vision of embodied intelligence necessitates agents that can operate beyond isolation [5, 37]. Realizing multi-agent systems capable of physical collaboration represents a critical next step. Yet, research in this domain remains disproportionately focused on single-robot tasks.

Despite growing interest in multi-agent interaction, current solutions are largely confined to virtual environments or simplified humanoid-object interactions. In computer graphics, physics-based animation has achieved realistic simulations of multi-character interactions [10, 25, 45, 50, 54], yet these methods often prioritize visual fidelity over the strict physical constraints necessary for real-world robotic deployment.

In robotics, although human–robot collaboration [9, 20] and competitive sports in structured environments [28, 39, 47] have been explored, these typically involve a compliant human partner or passive objects. Research explicitly targeting multi-humanoid interaction remains scarce, and existing works are predominantly validated only in simulation [29]. Consequently, achieving robust, physically coupled whole-body control on real multi-humanoid hardware remains an unbridged gap in the field.

Two fundamental challenges hinder the Sim-to-Real transition for dual-humanoid systems: (1) the scarcity of feasible interaction data; and (2) the complexity of the training-to-deployment paradigm. First, acquiring high-quality interaction references is non-trivial. Directly transferring *Human-Human Interaction* [38, 41, 45, 46] data to robots introduces severe kinematic conflicts due to morphological differences between humans and humanoids (see Sec. III-A). Standard retargeting methods [3, 30, 49] struggle to preserve both individual motion style and precise interaction geometry, yielding suboptimal motion references. Second, the learning and deployment pipeline presents significant hurdles. Existing tracking policies [26, 44, 52] typically treat agents as isolated entities, failing to model the intricate coupled dynamics essential for close interaction. Meanwhile, a significant disparity exists between the global observability available in simulation and the asynchronous, ego-centric, partially observable reality of real hardware, making the deployment unstable.

To address these challenges, we introduce **Rhythm** (Interactive Whole-Body Control for Dual Humanoids), a unified framework designed to empower dual humanoids to execute complex, physically plausible interactions in real-world scenarios. The framework explicitly models interaction geometry and physical contact to achieve high-fidelity coupled behaviors. First, to resolve kinematic conflicts in reference generation, we introduce **Interaction-Aware Motion Retargeting (IAMR)**. By explicitly modeling interaction geometry and utilizing distance-aware dynamic weighting, this module adaptively balances self-motion fidelity with interaction geometry. The resulting geometrically consistent references serve as a crucial prior, laying the physical foundation for the subsequent training phase. Building upon this, we develop **Interaction-Guided Reinforcement Learning (IGRL)** to master the complex dual-agent dynamics. This module directly leverages the interaction geometry and contact preserved by IAMR through explicit graph-based rewards, guiding agents to learn synchronized and robust behaviors. Finally, to realize dual-humanoid interactive whole-body control in the real world, we implement a relative state estimation and inter-agent synchronization scheme, effectively bridging the gap between global simulation and ego-centric reality.

Fig. 1 illustrates the robustness of our framework in real-world scenarios, successfully handling tasks ranging from intensive contacts to synchronized long-horizon dancing. Our main contributions are summarized as follows:

- We propose **Rhythm**, the first unified framework for whole-body dual-humanoid interaction that achieves the first successful **robust transfer** of complex interactive behaviors to physical hardware.
- We develop **IAMR** to resolve kinematic conflicts and generate humanoid-humanoid interaction motion references. Furthermore, we release **MAGIC**, the Multi-Humanoid Geometric Interaction Dataset, offering paired raw and retargeted interaction data.
- We introduce **IGRL**, a multi-agent learning module that masters coupled interaction dynamics. By incorporating graph-based rewards, it enables agents to learn robust, physically consistent interactive behaviors.
- We conduct extensive experiments on Unitree G1 humanoids in simulation and on physical hardware. Quantitative and qualitative evaluations across diverse interaction tasks demonstrate the superior performance and robustness of our framework.

II. Related Work

A. Learning-Based Humanoid Control and Interaction

Learning-based whole-body control methods primarily leverage Reinforcement Learning (RL) to endow humanoid robots with diverse motor skills [2, 4, 15–19, 24, 33, 40, 48, 52, 53]. Existing research largely focuses on single-agent

motion tracking [17, 44, 53], progressing from mimicking specific references to general motion tracking frameworks [7, 13, 26, 52, 55]. Beyond isolation, recent works have addressed the interaction between humanoids and environmental factors in sports such as table tennis [39], badminton [28], and soccer [34, 47], as well as human-robot collaboration [6, 9, 20]. However, these scenarios typically model the partner (human or object) as a passive entity or an external disturbance, ignoring the complex coupled dynamics inherent in multi-agent systems. Achieving interactive whole-body control on dual-humanoid hardware requires moving beyond these assumptions to explicitly model the mutual physical influence and geometric topology between active agents—a capability that remains absent in current robotics literature.

B. Physics-Based Animation of Human-Human Interaction

While robotics research focuses on hardware feasibility, the computer graphics community has expanded the scope of humanoid animation from single-agent generation [30–32] to the complex domain of *Human-Human Interaction* [38, 41, 46]. Addressing the inevitable physical artifacts caused by sensor inaccuracies in existing datasets [11, 25, 45, 51], recent works [10, 43, 50, 54] employ RL within physics-based simulations to guarantee human-like behavior and physical plausibility. However, these methods prioritize visual fidelity over dynamic feasibility, often relaxing rigorous constraints essential for real robots. Policies trained in these idealized environments struggle with the Sim-to-Real gap. To bridge this gap, we propose **Rhythm**, which pioneers the implementation of dual humanoid interaction in real-world scenarios.

C. Motion Retargeting for Humanoid

Motion retargeting serves as a fundamental bridge for transferring human skills to humanoids. However, standard approaches like PHC [30] and GMR [3] typically treat agents in isolation, neglecting interaction behaviors. While OmniRetarget [49] advances this by employing *Interaction Meshes* [1, 56] to enforce human-object interaction constraints, its applicability is restricted to single-humanoid settings. Extending this to human-human interaction presents unique challenges. Recent attempts, such as PAIR [20] and Harmanoid [29], adopt a coupled formulation that intertwines self-motion with interaction constraints. Consequently, in the presence of inherent kinematic conflicts, they often compromise self-motion fidelity (e.g., causing unnatural distortions) to strictly enforce interaction geometry. To address this, we propose IAMR that effectively mitigates these kinematic conflicts to produce high-fidelity coupled motion references.

III. Methods

As illustrated in Fig. 2, **Rhythm** addresses the challenge of dual-humanoid interaction through three tightly integrated components: 1) **Interaction-Aware Motion Retargeting (IAMR)**, which synthesizes physically feasible priors by decoupling interaction geometry from self-motion fidelity (Sec. III-A); 2) **Interaction-Guided Reinforcement Learning (IGRL)**, which captures coupled dynamics via topology-aware rewards that enforce interaction consistency (Sec. III-B); and 3) **Real-World Deployment**, which overcomes the limitations of noise global observability and asynchronous execution to bridge the Sim-to-Real gap (Sec. III-C).

A. Interaction-Aware Motion Retargeting (IAMR)

1) Problem Formulation: Given a source motion sequence of two human demonstrators with different anthropometry, our goal is to synthesize kinematically feasible trajectories for two humanoids. The core challenge lies in synthesizing motions that simultaneously preserve the individual motion style and the dense interaction geometry.

Interaction Mesh and Laplacian Coordinates. To mathematically model the coupled multi-agent system, we adopt the volumetric Interaction Mesh formalism [1, 56]. We represent the system as a connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, comprising a vertex set $\mathcal{V}$ of the agents' key joints and an edge set $\mathcal{E}$ encoding their structural connections. To encode local geometric details, we utilize Laplacian coordinates. For a vertex p_i, the Laplacian operator $\mathcal{L}(\cdot)$ computes the deviation from the weighted average of its neighbors $\mathcal{N}(i)$:

$$\mathcal{L}(p_i) = p_i - \sum_{j \in \mathcal{N}(i)} c_{ij} p_j,$$

where c_{ij} denotes the normalized weights. In mesh-based motion retargeting, the objective is to find a target configuration q such that the local geometry of the retargeted vertices $p_i(q)$ matches the source reference. This is achieved by minimizing the deformation energy $\sum \|\mathcal{L}(p_i(q)) - \mathcal{L}(p_i^{src})\|^2$.

The Kinematic Conflict. While the standard formulation $\sum \|\mathcal{L}(p_i(q)) - \mathcal{L}(p_i^{src})\|^2$ is effective for single-agent editing, applying it to the transfer from heterogeneous humans to homogeneous robots creates a fundamental ambiguity in defining the source reference $\mathcal{V}^{src}$. Due to the embodiment mismatch, no single reference manifold can simultaneously satisfy both self-motion and interaction constraints:

- **Individual Manifold ($\mathcal{M}_{ind}$):** Constructed by scaling each human with individual ratios to match the robot's height. Using this as $\mathcal{V}^{src}$ preserves valid self-motion Laplacian coordinates but structurally disrupts the relative interaction geometry (e.g., causing "air handshakes").
- **Unified Manifold ($\mathcal{M}_{uni}$):** Constructed by applying a single global scale to the entire scene. Using this as $\mathcal{V}^{src}$ strictly preserves relative interaction edges but forces the robots to adopt kinematic constraints incompatible with their morphology (e.g., foot floating).

Topological Partitioning. To resolve this conflict, we propose to relax the monolithic structure of the standard Interaction Mesh. Instead of treating the system as a single deformable body, we explicitly partition $\mathcal{E}$ into two disjoint functional groups, as visualized in the left part of Fig. 2: (1) **Intra-Agent Edges ($\mathcal{E}_{self}$):** Edges connecting joints within a single robot, encoding the local self-motion topology. (2) **Inter-Agent Edges ($\mathcal{E}_{inter}$):** Edges connecting key joints between the two robots, encoding the relative interaction topology.

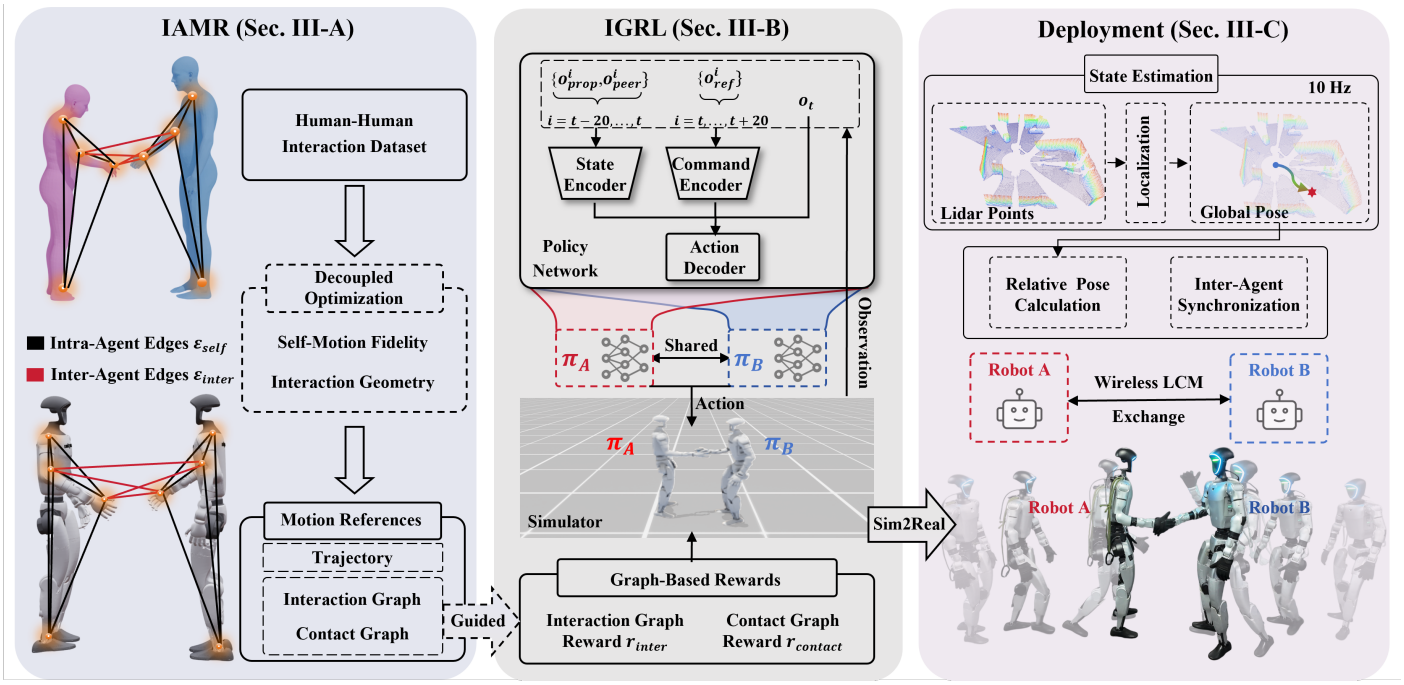

Fig. 2: Overview of Rhythm. IAMR utilizes decoupled optimization to generate high-quality humanoid-humanoid motion interaction references from human demonstrations. Guided by these references, IGRL employs MAPPO and graph-based rewards to learn robust coupled dynamics. Finally, the deployment module facilitates Sim-to-Real transfer via Lidar-fused state estimation and inter-agent synchronization.

This topological decomposition allows us to assign distinct geometric references to different semantic parts of the graph, forming the basis for our decoupled optimization scheme.

2) Decoupled Optimization: Leveraging the topological partitioning defined above, we resolve the kinematic conflict by assigning distinct geometric references to the partitioned subgraphs. We formulate the retargeting as a dynamic spring system, where a self-motion term E_{self} ensures intra-agent edges $\mathcal{E}_{self}$ track the Independent Manifold $\mathcal{M}_{ind}$, while an interaction term E_{inter} constrains inter-agent edges $\mathcal{E}_{inter}$ to the Unified Manifold $\mathcal{M}_{uni}$.

We solve for the optimal joint configuration $q^* = \{q_1, q_2\}$ by minimizing a hybrid energy function:

$$q^* = \arg\min_q \left(E_{self}(q) + E_{inter}(q)\right) \quad \text{s.t.} \quad q \in \mathcal{C}_{phy},$$

where $\mathcal{C}_{phy}$ represents the set of feasible configurations satisfying joint limits and collision constraints.

Self-Motion Objective (E_{self}). To preserve individual motion quality, we align the Laplacian geometry of each robot to its Independent Reference $\mathcal{M}_{ind}$, strictly confining the operator to the local subgraph:

$$E_{self}(q) = \sum_{a\in\{1,2\}} \sum_{p_i \in \mathcal{V}_a} \|\mathcal{L}(p_i) - \mathcal{L}(p_i^{ind})\|^2 + \lambda_{rot} \sum_{a\in\{1,2\}} \sum_{k\in\mathcal{B}_a} \|\theta_k \ominus \hat{\theta}_k^{src}\|^2,$$

where $\mathcal{V}_a$ denotes the vertex set of robot a, and p_i^{ind} represents the corresponding vertex position in the reference $\mathcal{M}_{ind}$. For rotation, $\mathcal{B}_a$ denotes the key links, where $\ominus$ measures the geodesic distance on $SO(3)$ between the current orientation θ_k and the reference $\hat{\theta}_k^{src}$.

Interaction Objective (E_{inter}). To enforce the relative interaction geometry, we treat the inter-agent edges as extrinsic constraints driven by the Unified Reference $\mathcal{M}_{uni}$. We formulate this as a variable-stiffness spring potential:

$$E_{inter}(q) = \sum_{(i,j)\in\mathcal{E}_{inter}} \omega_{ij}(d_{ij}) \cdot \|(p_i - p_j) - (\hat{p}_i^{uni} - \hat{p}_j^{uni})\|^2,$$

Here, p_i and p_j denote the current vertex positions of different agents, while $\hat{p}_i^{uni}$ and $\hat{p}_j^{uni}$ represent the corresponding target coordinates derived from the reference $\mathcal{M}_{uni}$. To naturally prioritize close-range geometry (e.g., contact) over distant relations, we define the stiffness ω_{ij} as a continuous exponential decay function of the source distance d_{ij}:

$$\omega_{ij}(d_{ij}) = \omega_{max} \cdot e^{-\gamma d_{ij}},$$

where ω_{max} denotes peak stiffness and γ the decay rate. This effectively models the interaction as a non-linear spring system that stiffens during close contact to prevent penetration, while becoming compliant at a distance to allow free motion.

Topological Interaction Priors. Beyond kinematic trajectories, IAMR explicitly extracts inter-agent topological structures to serve as interaction priors for downstream policy learning, as illustrated in the left part of Fig. 2. We generate two cross-agent graph representations: (1) An **Interaction Graph** (derived from $\mathcal{E}_{inter}$), which encodes the binary connectivity *bridging* the keypoints of the two agents [54]; (2) A **Contact Graph** (constructed via collision detection), which records the binary physical contact states *between* the links of the two distinct robots [42]. These graphs provide the essential interaction topology required for the graph-based rewards in the subsequent training phase.

B. Interaction-Guided Reinforcement Learning (IGRL)

To master the coupled dynamics of dual-humanoid interaction, we propose **IGRL**, a multi-agent reinforcement learning module. Unlike standard motion tracking policies that treat agents as isolated entities, IGRL explicitly models interaction geometry through graph-based rewards and incorporates specific training strategies to bridge the reality gap.

1) Multi-Agent Policy Design: As shown in the middle part of Fig. 2, we formulate the problem as a Multi-Agent Markov Decision Process (MA-MDP) and adopt the *Centralized Training with Decentralized Execution* (CTDE) paradigm using MAPPO [10, 27].

Interaction-Centric Observation. Effective collaboration requires a comprehensive awareness of both self and partner states. The policy operates on a compact observation space $o_t = \{o_{prop}, o_{peer}, o_{ref}\}$:

- Proprioception (o_{prop}): Defines the agent's internal state, including joint positions, velocities, base angular velocities, and the previous action.
- Peer Perception (o_{peer}): Encodes the peer's state relative to the ego agent. It includes the peer's joint positions and the relative root transform expressed in the ego-centric frame: the relative position $P_{rel} = R_{ego}^T(P_{peer} - P_{ego})$ and orientation $R_{rel} = R_{ego}^T R_{peer}$. This explicit spatial formation enables the agent to anticipate and handle coupled dynamics.
- Reference Motion (o_{ref}): Contains the future reference trajectories and the reference relative state, serving as the correct interaction topology.

Network Architecture. We design a hierarchical policy architecture where 1D-CNN temporal encoders process historical observations and future references, feeding latent features along with the current observation into an MLP action decoder. More details are provided in Appendix B.

Robust Training Strategy. To ensure transferability and handle complex interaction phases, we implement two strategies, with mathematical formulations provided in Appendix B:

- **Curriculum-based Adaptive Sampling:** Standard RSI [31] relies on sparse failure counts, neglecting non-terminal interaction violations. We propose an error-aware sampling strategy based on a multi-objective landscape composed of integrating failure, tracking, and interaction metrics. The curriculum dynamically evolves from prioritizing stability in early stages to focusing on tracking and interaction precision as the policy matures.
- **Dual-Agent Domain Randomization:** To bridge the sim-to-real gap, we simulate wireless latency via noisy, delayed peer observations and apply initial state perturbations to enforce recovery from physical misalignment.

2) Graph-based Rewards: Standard tracking rewards typically treat agents as isolated entities, failing to enforce topological consistency or capture coupled dynamics. To bridge this gap, we introduce graph-based rewards to **guide** the learning of dynamic control policies by translating the *Topological Interaction Priors* established in IAMR, as shown in Fig. 2.

Interaction Graph Reward (r_{inter}). To ensure precise relative geometry during interaction, we penalize deviations of the *interaction graph*. The reward is formulated as:

$$r_{inter} = \exp\left(-\frac{1}{\sigma_{inter}} \sum_{(i,j)\in\mathcal{E}_{inter}} \omega_{ij} \|d_{ij}^{sim} - \hat{d}_{ij}^{ref}\|^2\right),$$

where σ_{inter} is a sensitivity scaling factor, and d_{ij} denotes the relative position vector connecting joints i and j in the simulation (sim) and reference (ref) states. By inheriting the distance-aware dynamic weights ω_{ij} from IAMR, the policy inherently learns to prioritize the same geometric constraints that were optimized during retargeting.

Contact Graph Reward ($r_{contact}$). Since kinematic references lack force information, we utilize *contact graph* to regularize physical interaction. This reward is designed with two goals: (1) *Contact Consistency*, which penalizes mismatches between the simulated and reference contact states; and (2) *Force Regularization*, which constrains contact forces to realistic ranges when active, and penalizes non-zero forces during non-contact phases. This explicitly discourages interpenetration and encourages compliant physical interaction.

C. Real-World Deployment

Deploying **Rhythm** on physical hardware faces the *Sim-to-Real Gap*, specifically the lack of global observability and the issue of asynchronous execution.

1) State Estimation and Relative Localization: Constructing the observation space for dual-humanoid interaction requires precise global and relative state information. We employ a robust localization system, utilizing *POINT-LIO* [14] for high-frequency local odometry and *GICP* [35] to register real-time point clouds against a pre-built map for drift-free global positioning. A Kalman Filter fuses these estimates to ensure robustness under highly dynamic motions.

Robots broadcast global poses $\{P, R\}$ via *LCM* [21]. The ego-agent transforms received data into its local frame to derive $\{P_{rel}, R_{rel}\}$, enabling real-time reconstruction of o_{peer} consistent with the simulation.

2) Inter-Agent Synchronization: We use the *motion phase* ϕ [31] as the continuous variable representing the temporal execution progress of the interaction policy. In distributed settings, inherent hardware clock drift inevitably causes the phases of the two agents to diverge over time.

To maintain temporal alignment, we implement a soft synchronization mechanism based on proportional feedback. Agents exchange their current phase ϕ via the wireless bridge. Upon receiving the peer's phase ϕ_{peer}, the ego agent dynamically modulates its phase progression rate $\dot{\phi}_{ego}$. Instead of a fixed increment ($\dot{\phi}_{base} = 1.0$), we apply a correction:

$$\dot{\phi}_{ego} = 1.0 + k(\phi_{peer} - \phi_{ego}),$$

where k is the synchronization gain. This compensates for temporal drift by modulating execution rate, ensuring smooth alignment without the discontinuities of hard phase resets.

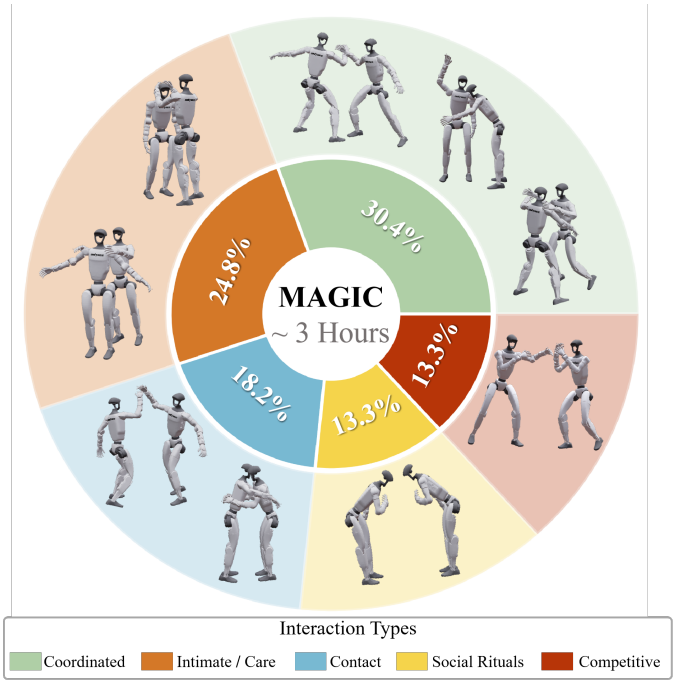

Fig. 3: Overview of MAGIC. MAGIC contains ∼3 hours of high-fidelity interaction data balanced across five semantic categories (inner chart). Representative snapshots (outer ring) illustrate the diversity ranging from loose spatiotemporal coordination to intensive contact.

IV. Experiments

We design our experiments to systematically validate the proposed framework by answering three core questions:

- **Q1 (Retargeting Quality):** Can IAMR synthesize kinematically feasible and topologically consistent trajectories that preserve the interaction geometry of the source data?
- **Q2 (Policy Efficacy):** Can IGRL utilize the interaction guidance to capture the coupled dynamics, overcoming the limitations of treating agents as isolated entities?
- **Q3 (Real-World Robustness):** Can the learned policies be successfully deployed to physical dual-humanoid systems, overcoming the limitations of noise global observability and asynchronous communication?

A. The MAGIC Dataset

High-quality motion data is the cornerstone of learning interactions. Addressing the lack of clean human-human interaction data, we introduce **MAGIC**, a high-fidelity motion capture dataset comprising ∼3 hours of valid motion sequences.

1) Acquisition and Diversity: Distinct from existing datasets [11, 25, 45, 51], we ensured physical plausibility for robot transfer through *Anthropometric Consistency* (matched actor heights) and *Temporal Continuity* (long-horizon sequences $>$ 10s). As illustrated in Fig. 3, MAGIC covers a diverse semantic spectrum: *Coordinated* actions (30.4%), *Intimate/Care* behaviors (24.8%), *Contact* (18.2%), *Social Rituals* (13.3%), and *Competitive* interactions (13.3%).

2) Data Release: To facilitate future research, we will publicly release both the raw motion capture data and the retargeted humanoid reference trajectories generated by IAMR.

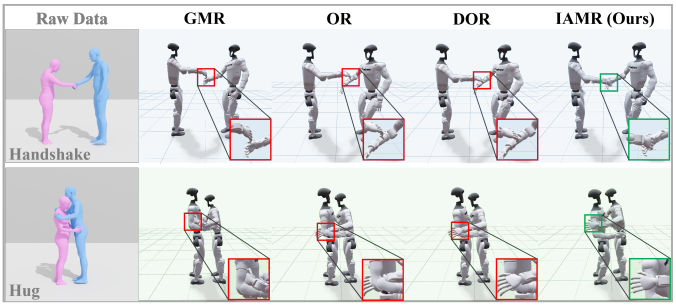

Fig. 4: Qualitative Visualization of Retargeting on Inter-X. Top: Baselines suffer from contact loss ("air handshakes"), whereas IAMR preserves precise interaction geometry. **Bottom:** OR leads to severe penetration while DOR forces unnatural stiff postures; IAMR maintains close-proximity topology without collisions.

B. Experimental Setup

Datasets. We use the proposed MAGIC dataset for both training and evaluation. To systematically assess performance across diverse physical dynamics, we categorize the tasks into three physics-based groups: **Collaborate** (non-contact synchronization), **Light Contact** (transient interaction), and **Intensive Contact** (continuous force transmission). Additionally, we employ the external *Inter-X* dataset [45] to assess robustness under significant anthropometric mismatches.

Baselines. We benchmark **Rhythm** against representative methods under two evaluation aspects, with full implementation details in Appendix C:

- **Retargeting Quality (Q1):** We compare against **GMR** [3], which performs Cartesian optimization, **OR** [49], a single-agent interaction mesh method, and **DOR**, our constructed dual-agent retargeting baseline.
- **Policy Efficacy (Q2):** To validate the effectiveness of IGRL, we compare against a **Single-Agent** (Status Quo) baseline that performs isolated tracking, as well as key ablated variants of our method, including **w/o Peer Obs**, **w/o Contact Rew**, and **w/o Interaction Rew**.

Evaluation Metrics. We adopt task-specific evaluation metrics for **Retargeting Quality (Q1)** and **Policy Efficacy (Q2)**. For conciseness, detailed metric definitions and mathematical formulations are deferred to Appendix C.

C. Retargeting Quality (Q1)

We evaluate IAMR against baselines across three dimensions: **Safety**, measured by Inter-Penetration Rate (IPR) and Max Penetration Depth (MPD); **Fidelity**, quantified by Interaction Edge Error (IEE) and Contact F1 Score [20, 29]; and **Utility**, assessed via Downstream Success Rate (DSR).

Quantitative Analysis. Table I presents the performance comparison across four interaction categories. We observe three key trends. First, isolated baselines (GMR, OR) fundamentally fail to ensure physical feasibility. Notably, in Intensive Contact, OR exhibits a prohibitive Inter-Penetration Rate (IPR) of 47.3%, rendering the motions physically infeasible and unsuitable for policy learning. Second, while the coupled baseline (DOR) guarantees safety (IPR=0), its rigid formulation compromises interaction fidelity under anthropometric

TABLE I: Quantitative Results of Retargeting. Comparison across four interaction categories. Metrics include Safety (IPR, MPD), Fidelity (IEE, F1), and Utility (DSR). IAMR achieves the best balance, strictly eliminating penetration (IPR=0) while maximizing contact F1 scores.

	Safety		Fidelity			Utility
	IPR(%)↓	MPD(cm)↓	IEE(%)↓	F1-S↑	F1-L↑	DSR(%)↑
MAGIC: Collaborate						
GMR	0.14	1.2	4.3	0.602	0.804	85.5
OR	0.20	1.4	4.1	0.747	0.902	87.4
DOR	**0.00**	**0.0**	3.9	0.711	0.899	**89.5**
IAMR	**0.00**	**0.0**	**3.7**	**0.785**	**0.936**	89.0
MAGIC: Light Contact						
GMR	2.18	3.3	4.6	0.738	0.893	48.5
OR	7.62	5.9	3.6	0.844	0.912	63.1
DOR	**0.00**	**0.0**	3.6	0.810	0.918	69.4
IAMR	**0.00**	**0.0**	**3.1**	**0.905**	**0.935**	**75.3**
MAGIC: Intensive Contact						
GMR	35.2	3.8	9.6	0.864	0.928	45.5
OR	47.3	5.3	8.0	0.884	0.929	56.5
DOR	**0.00**	**0.0**	7.8	0.883	0.925	63.3
IAMR	**0.00**	**0.0**	**6.6**	**0.932**	**0.941**	**78.3**
Inter-X						
GMR	11.7	1.7	8.0	0.598	0.752	31.7
OR	18.4	2.6	6.8	0.587	0.791	46.3
DOR	**0.00**	**0.0**	6.7	0.589	0.795	52.9
IAMR	**0.00**	**0.0**	**4.9**	**0.843**	**0.860**	**69.9**

mismatch. This is evident on the Inter-X dataset, where DOR's F1-Strict drops to 0.589 due to its inability to reconcile conflicting kinematic constraints. In contrast, IAMR achieves the optimal safety-fidelity trade-off. By adaptively decoupling self-motion from interaction objectives, our method eliminates penetration while outperforming DOR by 43% in F1-Strict on Inter-X. Crucially, this strict preservation of interaction topology translates to superior downstream utility, yielding 78.3% DSR in contact-rich tasks.

Qualitative Visualization (Inter-X Cases). Fig. 4 validates robustness under significant mismatches. In transient "Handshakes", IAMR utilizes dynamic weighting to rioritize critical interaction geometry, effectively resolving the contact loss ("air handshakes") observed in baselines. Conversely, in continuous "Hugs", our decoupled optimization reconciles kinematic conflicts, preventing the penetration of OR and the stiffness of DOR to maintain valid close-proximity topology.

D. Policy Efficacy (Q2)

We assess the robustness and fidelity of the learned control policy across two dimensions: **Interaction Performance**, measured by Interaction Edge Error (IEE) and Interaction Success Rate (ISR); and **Contact Performance**, quantified by Contact Success Rate (CSR) and Contact Error Rate (CER).

Quantitative Analysis. Table II presents the quantitative ablation results, revealing two critical insights. First, Interaction Awareness is non-negotiable. The Single Agent (Vanilla) baseline, limiting agents to isolated tracking, fails to coordinate effectively, achieving only 18.7% ISR in Collaborate scenarios. Similarly, removing peer observations (w/o Peer Obs) severs the closed-loop synchronization, causing physical coupling to collapse (CSR drops to 18.6% in Light Contact). Second, there exists a functional hierarchy between coarse-grained geometric guidance and fine-grained contact regulation. The ablation results reveal that the interaction reward serves as a necessary foundation: removing it (w/o Interaction Rew) causes performance to collapse across all metrics as the policy fails to guide agents into the interaction envelope where contact can be established. For example, ISR drops to 58.1% in Collaborate and CSR drops to 28.1% in Light Contact. Once this spatial proximity is achieved, the contact reward becomes critical for physical realism. Notably, the w/o Contact Rew variant exhibits a "ghosting" phenomenon: it attains high geometric precision (93.4% ISR in Collaborate) by disregarding physical collision constraints, yet fails to maintain valid physical contact (only 52.1% CSR in Light Contact). Ours (Full) effectively integrates these components, leveraging geometric guidance to establish the spatial foundation while using contact regulation to enforce valid physical coupling (above 75% ISR and 77% CSR across all scenarios).

TABLE II: Quantitative Results of Policy. We evaluate the contribution of each component. Our full method achieves the most robust balance, effectively integrating coarse-grained geometric alignment (low IEE) with fine-grained physical contact fidelity (high CSR).

	Interaction		Contact	
	ISR(%)↑	IEE(%)↓	CSR(%)↑	CER↓
MAGIC: Collaborate				
Single Agent	18.7	38.9	100.0	0.000
w/o Peer Obs	19.5	47.0	100.0	0.000
w/o Contact Rew	**93.4**	**4.7**	100.0	0.000
w/o Interact Rew	58.1	15.1	100.0	0.000
Ours (Full)	92.9	4.8	100.0	0.000
MAGIC: Light Contact				
Single Agent	34.3	19.9	24.1	0.283
w/o Peer Obs	48.9	13.9	18.6	0.268
w/o Contact Rew	85.9	5.4	52.1	0.203
w/o Interact Rew	48.7	19.3	28.1	0.243
Ours (Full)	**90.0**	**4.2**	**78.0**	**0.120**
MAGIC: Intensive Contact				
Single Agent	24.0	29.9	37.5	0.312
w/o Peer Obs	34.1	21.4	43.7	0.280
w/o Contact Rew	**77.3**	**7.7**	70.6	0.174
w/o Interact Rew	51.3	17.0	56.8	0.211
Ours (Full)	75.2	7.9	**78.8**	**0.159**
Inter-X				
Single Agent	25.7	26.9	57.4	0.256
w/o Peer Obs	73.2	7.6	75.3	0.143
w/o Contact Rew	**95.1**	**3.4**	68.3	0.208
w/o Interact Rew	63.2	9.7	**78.6**	**0.110**
Ours (Full)	92.8	3.5	77.4	0.125

Qualitative Visualization. Fig. 5 highlights the behavioral divergence across methods in the Greeting task. The Single Agent, treating the peer merely as a dynamic obstacle, may

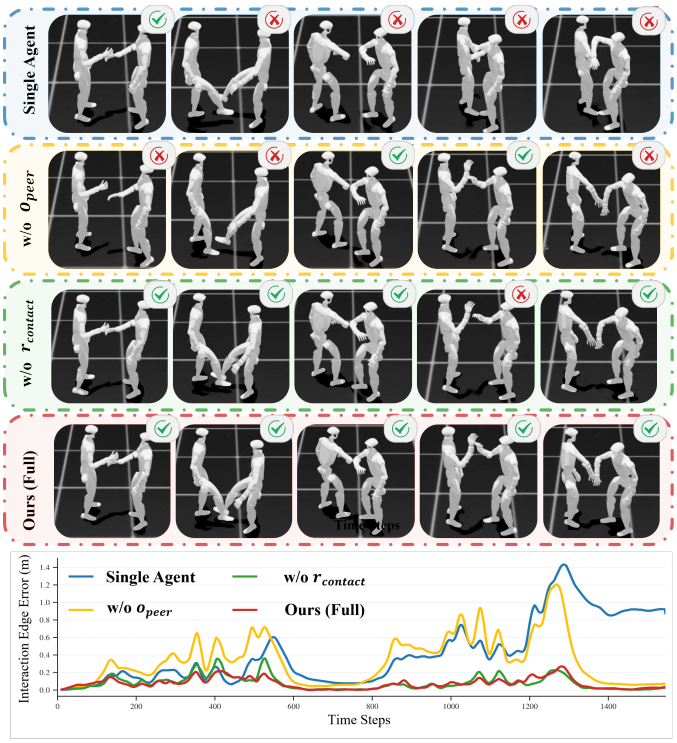

Fig. 5: Qualitative Visualization of Policy. Single Agent (blue) drifts into collisions. w/o Contact Rew (green) achieves low error but exhibits physical "ghosting". In contrast, Ours enforces valid physical contact.

initiate interaction but inevitably drifts into collision. While its low-level robustness prevents immediate termination, the interaction topology is destroyed. Similarly, without explicit peer monitoring, the w/o Peer Obs baseline suffers from severe desynchronization (large error spikes). Notably, the w/o Contact Rew variant successfully maintains coarse-grained geometric alignment, yielding an IEE curve comparable to the full method (Green versus Red lines). However, it lacks fine-grained contact fidelity, allowing hands to "ghost" through each other. Ours bridges this gap, achieving the precision as w/o Contact Rew while enforcing valid physical coupling at the contact interface. These visual observations directly align with the quantitative hierarchy in Table II, particularly explaining the discrepancy between geometric (ISR) and physical (CSR) metrics in the baselines.

E. Real-World Robustness (Q3)

We validate the deployment of **Rhythm** on Unitree G1 humanoids to assess its generality, robustness, and success rate in physical environments.

Framework Generality. Fig. 1 demonstrates the system's versatility across diverse modalities. By strictly preserving fine-grained contact geometry in physical coupling tasks (Fig. 1a-c) and maintaining spatiotemporal coherence in long-horizon coordination (Fig. 1d), **Rhythm** effectively ensures interaction integrity across the spectrum using a unified formulation.

Quantitative Success Rate. Table III reports success rates (measured over 10 trials) based on valid contact establishment. **Rhythm** consistently outperforms the Single Agent baseline by more than 60%, with the most pronounced gap in *Greeting* (12.2% versus 82.2%). While the baseline fails due to accumulated drift in this long-horizon task, our relative state estimation continuously compensates for misalignment, ensuring robust physical coupling.

TABLE III: Main Results for Real Robot Experiments. We conducted 10 trials for each task and evaluated success based on contact establishment at specific keyframes (K frames per trial).

Task	Method	Success / Total	Rate (%)
Hug	Single Agent	8 / 30	26.7%
	Ours	**26 / 30**	**86.7%**
Shoulder	Single Agent	6 / 30	20.0%
	Ours	**24 / 30**	**80.0%**
Greeting	Single Agent	11 / 90	12.2%
	Ours	**74 / 90**	**82.2%**

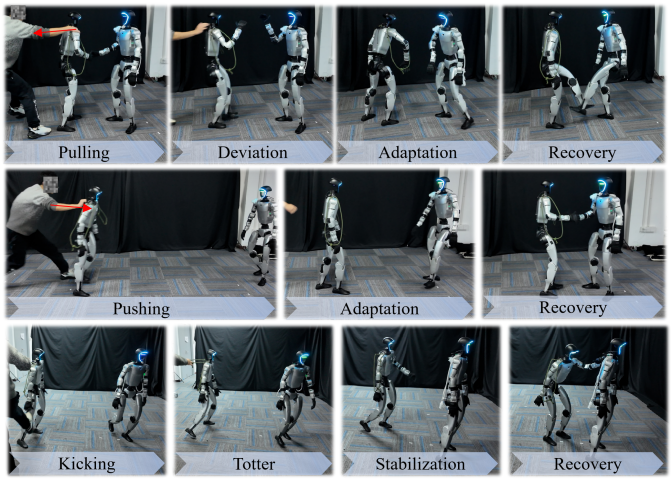

Fig. 6: Robustness to disturbances. Our policy demonstrates strong resilience against aggressive external perturbations (pulling, pushing, and kicking), successfully recovering balance and synchronization.

Robustness to Disturbances. To evaluate system resilience, we subjected the robots to significant external perturbations during execution. As illustrated in Fig. 6, these disturbances included aggressive pushing, pulling, and kicking forces. Despite the severity of these physical interferences, the agents successfully recovered their balance and actively adjusted their relative states to restore the interaction topology. This empirical evidence confirms that our policy has learned robust closed-loop synchronization capabilities, allowing for real-time recovery strategies rather than open-loop motion replay.

V. Conclusion

In this work, we present **Rhythm**, a unified framework that achieves the first robust transfer of physically coupled interactive behaviors to dual-humanoid hardware. By integrating Interaction-Aware Motion Retargeting (IAMR) to resolve kinematic conflicts and Interaction-Guided Reinforcement Learning (IGRL) to master coupled dynamics, our approach effectively bridges the Sim-to-Real gap. Furthermore, we release the MAGIC dataset to facilitate future research in multi-agent embodied intelligence.

While this work utilizes dual-humanoid setups to rigorously validate the modeling of coupled dynamics, our graph-based formulation is theoretically generic and supports extension to multi-agent systems. A current limitation is the reliance on pre-built maps for state estimation. Future work will focus on eliminating this dependency by shifting towards fully ego-centric perception for map-free collaboration, and scaling the framework to orchestrate complex multi-humanoid interactions in open-world scenarios.

REFERENCES

[1] Marc Alexa. Differential coordinates for local mesh morphing and deformation. *The Visual Computer*, 19 (2):105–114, 2003.

[2] Arthur Allshire, Hongsuk Choi, Junyi Zhang, David McAllister, Anthony Zhang, Chung Min Kim, Trevor Darrell, Pieter Abbeel, Jitendra Malik, and Angjoo Kanazawa. Visual imitation enables contextual humanoid control. In *Conference on Robot Learning*, 2025.

[3] Joao Pedro Araujo, Yanjie Ze, Pei Xu, Jiajun Wu, and C Karen Liu. Retargeting Matters: General motion retargeting for humanoid motion tracking. *arXiv preprint arXiv:2510.02252*, 2025.

[4] Qingwei Ben, Feiyu Jia, Jia Zeng, Junting Dong, Dahua Lin, and Jiangmiao Pang. HOMIE: Humanoid loco-manipulation with isomorphic exoskeleton cockpit. In *Robotics: Science and Systems*, 2025.

[5] Longbing Cao. Humanoid robots and humanoid ai: Review, perspectives and directions. *ACM Computing Surveys*, 58(4):1–37, 2025.

[6] Haoran Chen, Yiteng Xu, Yiming Ren, Yaoqin Ye, Xinran Li, Ning Ding, Yuxuan Wu, Yaoze Liu, Peishan Cong, Ziyi Wang, et al. Symbridge: A human-in-the-loop cyber-physical interactive system for adaptive human-robot symbiosis. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*, pages 1–12, 2025.

[7] Zixuan Chen, Mazeyu Ji, Xuxin Cheng, Xuanbin Peng, Xue Bin Peng, and Xiaolong Wang. GMT: General motion tracking for humanoid whole-body control. *arXiv preprint arXiv:2506.14770*, 2025.

[8] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. In *Robotics: Science and Systems*, 2024.

[9] Yushi Du, Yixuan Li, Baoxiong Jia, Yutang Lin, Pei Zhou, Wei Liang, Yanchao Yang, and Siyuan Huang. Learning human-humanoid coordination for collaborative object carrying. *arXiv preprint arXiv:2510.14293*, 2025.

[10] Jiawei Gao, Ziqin Wang, Zeqi Xiao, Jingbo Wang, Tai Wang, Jinkun Cao, Xiaolin Hu, Si Liu, Jifeng Dai, and Jiangmiao Pang. CooHOI: Learning cooperative human-object interaction with manipulated object dynamics. *Advances in Neural Information Processing Systems*, 37: 79741–79763, 2024.

[11] Anindita Ghosh, Rishabh Dabral, Vladislav Golyanik, Christian Theobalt, and Philipp Slusallek. ReMoS: 3D motion-conditioned reaction synthesis for two-person interactions. In *European Conference on Computer Vision (ECCV)*, 2024.

[12] Xinyang Gu, Yen-Jen Wang, Xiang Zhu, Chengming Shi, Yanjiang Guo, Yichen Liu, and Jianyu Chen. Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning. In *Robotics: Science and Systems*, 2024.

[13] Jinrui Han, Weiji Xie, Jiakun Zheng, Jiyuan Shi, Weinan Zhang, Ting Xiao, and Chenjia Bai. KungfuBot2: Learning versatile motion skills for humanoid whole-body control. *arXiv preprint arXiv:2509.16638*, 2025.

[14] Dongjiao He, Wei Xu, Nan Chen, Fanze Kong, Chongjian Yuan, and Fu Zhang. Point-LIO: robust high-bandwidth light detection and ranging inertial odometry. *Advanced Intelligent Systems*, 5(7):2200459, 2023.

[15] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. OmniH2O: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. In *Conference on Robot Learning*, 2024.

[16] Tairan He, Zhengyi Luo, Wenli Xiao, Chong Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Learning human-to-humanoid real-time whole-body teleoperation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2024.

[17] Tairan He, Jiawei Gao, Wenli Xiao, Yuanhang Zhang, Zi Wang, Jiashun Wang, Zhengyi Luo, Guanqi He, Nikhil Sobanbabu, Chaoyi Pan, Zeji Yi, Guannan Qu, Kris Kitani, Jessica K. Hodgins, Linxi Fan, Yuke Zhu, Changliu Liu, and Guanya Shi. ASAP: Aligning simulation and real-world physics for learning agile humanoid whole-body skills. In *Robotics: Science and Systems*, 2025.

[18] Xialin He, Runpei Dong, Zixuan Chen, and Saurabh Gupta. Learning getting-up policies for real-world humanoid robots. In *Robotics: Science and Systems*, 2025.

[19] Tao Huang, Junli Ren, Huayi Wang, Zirui Wang, Qingwei Ben, Muning Wen, Xiao Chen, Jianan Li, and Jiangmiao Pang. Learning humanoid standing-up control across diverse postures. In *Robotics: Science and Systems*, 2025.

[20] Wei-Jin Huang, Yue-Yi Zhang, Yi-Lin Wei, Zhi-Wei Xia, Juantao Tan, Yuan-Ming Li, Zhilin Zhao, and Wei-Shi Zheng. Learning whole-body human-humanoid interaction from human-human demonstrations. *arXiv preprint arXiv:2601.09518*, 2026.

[21] Yih Huang, Eric Fleury, and Philip K McKinley. LCM: A multicast core management protocol for link-state routing networks. In *ICC'98. 1998 IEEE International Conference on Communications. Conference Record. Affiliated with SUPERCOMM'98 (Cat. No. 98CH36220)*, volume 2, pages 1197–1201. IEEE, 1998.

[22] Jialong Li, Xuxin Cheng, Tianshu Huang, Shiqi Yang, Ri-Zhao Qiu, and Xiaolong Wang. AMO: Adaptive motion optimization for hyper-dexterous humanoid whole-body control. In *Robotics: Science and Systems*, 2025.

[23] Junheng Li, Ziwei Duan, Junchao Ma, and Quan Nguyen. Gait-Net-augmented implicit kino-dynamic MPC for dynamic variable-frequency humanoid locomotion over discrete terrains. In *Robotics: Science and Systems*, 2025.

[24] Yixuan Li, Yutang Lin, Jieming Cui, Tengyu Liu, Wei Liang, Yixin Zhu, and Siyuan Huang. CLONE: Closed-loop whole-body humanoid teleoperation for long-horizon tasks. In *Conference on Robot Learning*, 2025.

[25] Han Liang, Wenqian Zhang, Wenxuan Li, Jingyi Yu, and Lan Xu. InterGen: Diffusion-based multi-human motion generation under complex interactions. *International Journal of Computer Vision (IJCV)*, 2024.

[26] Qiayuan Liao, Takara E Truong, Xiaoyu Huang, Yuman Gao, Guy Tevet, Koushil Sreenath, and C Karen Liu. Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion. *arXiv preprint arXiv:2508.08241*, 2025.

[27] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the Eleventh International Conference on Machine Learning (ICML)*, volume 157, pages 157–163. Morgan Kaufmann, 1994.

[28] Chenhao Liu, Leyun Jiang, Yibo Wang, Kairan Yao, Jinchen Fu, and Xiaoyu Ren. Humanoid whole-body badminton via multi-stage reinforcement learning. *arXiv preprint arXiv:2511.11218*, 2025.

[29] Zuhong Liu, Junhao Ge, Minhao Xiong, Jiahao Gu, Bowei Tang, Wei Jing, and Siheng Chen. It Takes Two: Learning interactive whole-body control between humanoid robots. *arXiv preprint arXiv:2510.10206*, 2025.

[30] Zhengyi Luo, Jinkun Cao, Alexander Winkler, Kris Kitani, and Weipeng Xu. Perpetual humanoid control for real-time simulated avatars. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10895–10904, 2023.

[31] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14, 2018.

[32] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. ASE: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions on Graphics (TOG)*, 41 (4):1–17, 2022.

[33] Skand Peri, Akhil Perincherry, Bikram Pandit, and Stefan Lee. Non-conflicting energy minimization in reinforcement learning based robot control. In *Conference on Robot Learning*, 2025.

[34] Junli Ren, Junfeng Long, Tao Huang, Huayi Wang, Zirui Wang, Feiyu Jia, Wentao Zhang, Jingbo Wang, Ping Luo, and Jiangmiao Pang. Humanoid Goalkeeper: Learning from position conditioned task-motion constraints. *arXiv preprint arXiv:2510.18002*, 2025.

[35] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-ICP. In *Robotics: Science and Systems*, volume 2, page 435. Seattle, WA, 2009.

[36] Yiyang Shao, Bike Zhang, Qiayuan Liao, Xiaoyu Huang, Yuman Gao, Yufeng Chi, Zhongyu Li, Sophia Shao, and Koushil Sreenath. LangWBC: Language-directed humanoid whole-body control via end-to-end learning. In *Robotics: Science and Systems*, 2025.

[37] Qincheng Sheng, Zhongxiang Zhou, Jinhao Li, Xiangyu Mi, Pingyu Xiang, Zhenghan Chen, Haocheng Xu, Shenhan Jia, Xiyang Wu, Yuxiang Cui, et al. A comprehensive review of humanoid robots. *SmartBot*, 1(1):e12008, 2025.

[38] Sebastian Starke, Yiwei Zhao, Taku Komura, and Kazi Zaman. Local motion phases for learning multi-contact character movements. *ACM Transactions on Graphics*, 39(4), 2020.

[39] Zhi Su, Bike Zhang, Nima Rahmanian, Yuman Gao, Qiayuan Liao, Caitlin Regan, Koushil Sreenath, and S Shankar Sastry. HITTER: A humanoid table tennis robot via hierarchical planning and learning. *arXiv preprint arXiv:2508.21043*, 2025.

[40] Huayi Wang, Zirui Wang, Junli Ren, Qingwei Ben, Tao Huang, Weinan Zhang, and Jiangmiao Pang. BeamDojo: Learning agile humanoid locomotion on sparse footholds. In *Robotics: Science and Systems*, 2025.

[41] Lipeng Wang, Hongxing Fan, Haohua Chen, Zehuan Huang, and Lu Sheng. InterMoE: Individual-specific 3d human interaction generation via dynamic temporal-selective moe. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026.

[42] Yinhuai Wang, Qihan Zhao, Runyi Yu, Hok Wai Tsui, Ailing Zeng, Jing Lin, Zhengyi Luo, Jiwen Yu, Xiu Li, Qifeng Chen, Jian Zhang, Lei Zhang, and Ping Tan. Skillmimic: Learning basketball interaction skills from demonstrations. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17540–17549, June 2025.

[43] Jungdam Won, Deepak Gopinath, and Jessica Hodgins. Control strategies for physically simulated characters performing two-player competitive sports. *ACM Transactions on Graphics (TOG)*, 40(4):1–11, 2021.

[44] Weiji Xie, Jinrui Han, Jiakun Zheng, Huanyu Li, Xinzhe Liu, Jiyuan Shi, Weinan Zhang, Chenjia Bai, and Xuelong Li. KungfuBot: Physics-based humanoid whole-body control for learning highly-dynamic skills. *Advances in Neural Information Processing Systems*, 2025.

[45] Liang Xu, Xintao Lv, Yichao Yan, Xin Jin, Shuwen Wu, Congsheng Xu, Yifan Liu, Yizhou Zhou, Fengyun Rao, Xingdong Sheng, Yunhui Liu, Wenjun Zeng, and Xiaokang Yang. Inter-X: Towards versatile human-human interaction analysis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[46] Liang Xu, Chengqun Yang, Zili Lin, Fei Xu, Yifan Liu, Congsheng Xu, Yiyi Zhang, Jie Qin, Xingdong Sheng, Yunhui Liu, et al. Perceiving and acting in first-person:

A dataset and benchmark for egocentric human-object-human interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12535–12548, 2025.

[47] Zifan Xu, Myoungkyu Seo, Dongmyeong Lee, Hao Fu, Jiaheng Hu, Jiaxun Cui, Yuqian Jiang, Zhihan Wang, Anastasiia Brund, Joydeep Biswas, et al. Learning agile striker skills for humanoid soccer robots from noisy sensory input. *arXiv preprint arXiv:2512.06571*, 2025.

[48] Yufei Xue, Wentao Dong, Minghuan Liu, Weinan Zhang, and Jiangmiao Pang. A unified and general humanoid whole-body controller for fine-grained locomotion. In *Robotics: Science and Systems*, 2025.

[49] Lujie Yang, Xiaoyu Huang, Zhen Wu, Angjoo Kanazawa, Pieter Abbeel, Carmelo Sferrazza, C Karen Liu, Rocky Duan, and Guanya Shi. OmniRetarget: Interaction-preserving data generation for humanoid whole-body loco-manipulation and scene interaction. *arXiv preprint arXiv:2509.26633*, 2025.

[50] Wei Yao, Yunlian Sun, Chang Liu, Hongwen Zhang, and Jinhui Tang. PhysiInter: Integrating physical mapping for high-fidelity human interaction generation. *arXiv preprint arXiv:2506.07456*, 2025.

[51] Yifei Yin, Chen Guo, Manuel Kaufmann, Juan Jose Zarate, Jie Song, and Otmar Hilliges. Hi4D: 4D instance segmentation of close human interaction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[52] Yanjie Ze, Zixuan Chen, Joao Pedro Araujo, Zi-ang Cao, Xue Bin Peng, Jiajun Wu, and Karen Liu. TWIST: Teleoperated whole-body imitation system. In *Conference on Robot Learning*, 2025.

[53] Tong Zhang, Boyuan Zheng, Ruiqian Nai, Yingdong Hu, Yen-Jen Wang, Geng Chen, Fanqi Lin, Jiongye Li, Chuye Hong, Koushil Sreenath, and Yang Gao. HuB: Learning extreme humanoid balance. In *Conference on Robot Learning*, 2025.

[54] Yunbo Zhang, Deepak Gopinath, Yuting Ye, Jessica Hodgins, Greg Turk, and Jungdam Won. Simulation and retargeting of complex multi-character interactions. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023.

[55] Zhikai Zhang, Jun Guo, Chao Chen, Jilong Wang, Chenghuai Lin, Yunrui Lian, Han Xue, Zhenrong Wang, Maoqi Liu, Jiangran Lyu, et al. Track any motions under any disturbances. *arXiv preprint arXiv:2509.13833*, 2025.

[56] Kun Zhou, Jin Huang, John Snyder, Xinguo Liu, Hujun Bao, Baining Guo, and Heung-Yeung Shum. Large mesh deformation using the volumetric graph laplacian. *ACM Transactions on Graphics (TOG)*, 24(3):496–503, 2005.

APPENDIX

Overview

This appendix is organized into three main sections (A–C) to support the clarity and reproducibility of the proposed framework, **Rhythm**.

Terminology. We refer to our unified framework as **Rhythm**. Within this framework, we define two core components:

- **IAMR** (Interaction-Aware Motion Retargeting): The retargeting module that resolves kinematic conflicts to generate geometrically consistent motion references from heterogeneous human data (Sec. A).
- **IGRL** (Interaction-Guided Reinforcement Learning): The multi-agent learning module that masters coupled dynamics via graph-based rewards (Sec. B).

Structure. The Appendix is organized as follows:

- **A. Details of IAMR (Sec. A):** We first elaborate on the data processing pipeline for heterogeneous motion sources (Part 1). We then provide the mathematical formulations and constraints for the optimization problem (Part 2), followed by a visualization of the topological interaction priors (Part 3).
- **B. Details of IGRL (Sec. B):** This section specifies the hierarchical network architecture (Part 1), presents the comprehensive definition of graph-based rewards (Part 2), and details the robust training strategies, including curriculum learning and domain randomization (Part 3).
- **C. Experimental Setup & Metrics (Sec. C):** We provide implementation details for the benchmarking baselines (Part 1), followed by the rigorous mathematical definitions of the evaluation metrics (Part 2). Finally, we describe the hardware configuration and localization system used for Sim-to-Real transfer (Part 3).

A. Details of IAMR

1) Compatibility with Heterogeneous Motion Sources: Human motion datasets contain rich pose and interaction information but differ significantly in data format and physical attributes (e.g., height, body proportions). A key strength of our framework is its input-agnostic design: we first abstract diverse inputs into a standardized representation—time-series of *raw* global 3D keypoint positions $\{p_{t,i}^{\text{raw}}\}$—and then process them into two distinct reference manifolds to resolve the kinematic conflict.

Standardization of Input Formats. Regardless of the source format, our first step is to extract the raw 3D keypoints $p_{t,i}^{\text{raw}}$ for each agent $k \in \{1, 2\}$:

- **Parametric Human Models (SMPL):** The **Inter-X** dataset [45] utilizes the SMPL format. We compute the raw keypoints via the SMPL forward kinematics function $M(\cdot)$ using the recorded pose q and shape β:

$$p_t^{\text{raw},(k)} = M(q_t^{(k)}; \beta^{(k)}).$$

- **Skeleton Hierarchy (BVH):** Our **MAGIC** dataset utilizes the standard BVH skeleton hierarchy. Here, raw keypoints are derived directly from the skeleton's forward kinematics $f^{\text{skel}}(\cdot)$:

$$p_t^{\text{raw},(k)} = f^{\text{skel}}(q_t^{(k)}).$$

Construction of Dual Reference Manifolds. As discussed in the *Kinematic Conflict* (Sec. III-A), directly using raw human keypoints is infeasible due to morphological mismatches. To decouple interaction semantics from individual embodiment, we generate two distinct sets of scaled reference keypoints, P_{ind} and P_{uni}, corresponding to the individual ($\mathcal{M}_{ind}$) and unified ($\mathcal{M}_{uni}$) manifolds respectively.

Let h_{robot} be the robot's height, and $h_{\text{raw}}^{(k)}$ be the height of the k-th human demonstrator derived from $p^{\text{raw},(k)}$. We first compute the individual height ratio $s^{(k)} = h_{\text{robot}}/h_{\text{raw}}^{(k)}$.

- **Individual Scaling for Self-Motion ($\mathcal{M}_{ind}$):** To ensure kinematic feasibility for each robot's self-motion (e.g., limb proportions, ground contact), we generate the individual reference set P_{ind} by scaling each agent's raw keypoints with its own specific ratio $s^{(k)}$:

$$p_{t,i}^{\text{ind},(k)} = s^{(k)} \cdot p_{t,i}^{\text{raw},(k)}.$$

  This set P_{ind} serves as the target for all single-agent tracking objectives (e.g., $\mathcal{L}_{pos}$, $\mathcal{L}_{reg}$), ensuring that each robot tracks a trajectory compatible with its own scale.

- **Unified Scaling for Interaction Geometry ($\mathcal{M}_{uni}$):** To preserve the relative interaction topology (e.g., hand-holding distance), we generate the unified reference set P_{uni} by applying a single global scale $s_{unified}$ to both agents. We define this scale as the average of the individual factors:

$$s_{unified} = \frac{s^{(1)} + s^{(2)}}{2}, \quad p_{t,i}^{\text{uni},(k)} = s_{unified} \cdot p_{t,i}^{\text{raw},(k)}.$$

  This set P_{uni} is exclusively used to compute the *Interaction Graph* targets (e.g., relative edge lengths). This ensures that the spatial relationship between agents remains consistent with the original performance, preventing the geometric distortion that would arise from non-uniform scaling.

2) Optimization Details and Hyperparameters: We explicitly formulate the optimization objectives and constraints used to solve the kinematic conflict described in Sec. III-A.

Optimization Formulation. We solve the retargeting problem frame-by-frame using a Sequential Quadratic Programming (SQP) approach. For each frame t, we optimize the joint configurations $q_t = [q_t^{(1)}, q_t^{(2)}]$ to minimize the following total objective:

$$\mathcal{J}(q_t) = w_{self}\mathcal{J}_{self} + w_{inter}\mathcal{J}_{inter} + w_{reg}\mathcal{J}_{reg}.$$

The individual terms are defined as follows:

- **Self-Motion Preservation ($\mathcal{J}_{self}$):** To maintain local topology and orientation, we minimize deviations from

the *individual* reference manifold P_{ind}.

$$\mathcal{J}_{self} = \sum_{k\in\{1,2\}} \Big(\|\mathcal{L}(f(q_t^{(k)})) - \mathcal{L}(P_{ind}^{(k)})\|^2 + \lambda_{rot} \sum_{b\in\mathcal{B}_k} \|\theta_b(q_t^{(k)}) \ominus \hat{\theta}_b^{src}\|^2 \Big).$$

Here, $\mathcal{L}(\cdot)$ denotes the discrete Laplacian operator, and θ_b represents key bone orientations. This term allows the robot to adapt its absolute posture while preserving local motion semantics.

- **Interaction Preservation ($\mathcal{J}_{inter}$):** This term enforces relative geometric consistency by tracking the *unified* reference manifold P_{uni}. We split the interaction error term to fit the column width:

$$\mathcal{J}_{inter} = \sum_{(i,j)\in\mathcal{E}_{inter}} \omega_{ij} \big\| (f_i(q_t^{(1)}) - f_j(q_t^{(2)})) - (p_{t,i}^{uni,(1)} - p_{t,j}^{uni,(2)}) \big\|^2.$$

Where ω_{ij} is the distance-dependent stiffness. This explicitly penalizes deviations in the relative position vectors between agents.

- **Regularization ($\mathcal{J}_{reg}$):** Ensures temporal smoothness. It includes a minimum velocity term $\|q_t - q_{t-1}\|^2$.

Constraints. The optimization is subject to the following hard constraints:

- **Joint Limits:** The solution must respect the robot's physical joint ranges: $q^{min} \le q_{t-1} + \Delta q_t \le q^{max}$.
- **Collision Avoidance:** We enforce non-penetration for all active collision pairs. Linearizing the signed distance field $\phi(q)$, we require $J_{col}\Delta q_t \ge -\phi(q_{t-1}) - \epsilon_{safe}$, where J_{col} is the normal Jacobian and ϵ_{safe} is a safety margin.
- **Foot Contact:** For feet in strict contact (detected from source motion), we impose a zero-velocity constraint on the end-effectors: $\|J_{foot}\Delta q_t\| \le \epsilon_{stick}$.
- **Trust Region:** To ensure the validity of the linearization, we bound the step size: $\|\Delta q_t\|_2 \le \delta$.

Implementation & Hyperparameters. The optimization is implemented in Python using **CVXPY** with the **OSQP** solver. Empirically, we set the weights to prioritizes interaction stability: $w_{self} = 2.0$, $w_{inter} = 10.0$, and $w_{reg} = 0.1$. The rotation weight λ_{rot} is set to 0.1.

3) Topological Graph Visualization: To provide an intuitive understanding of the topological priors, we visualize the extracted graph structures using a representative interaction case (e.g., a handshake task), as shown in Fig. 7. The visualization highlights two distinct connectivity types used by IAMR:

- **The Interaction Graph (Yellow Edges):** Bridges the keypoints of the two agents based on the Unified Manifold. It captures the *spatial intent* and proximity required for coordination.
- **The Contact Graph (Red Edges):** Highlights active physical collision links between specific robot bodies. This explicitly encodes the *physical coupling* state.

Algorithm 1 Interaction-Aware Motion Retargeting (IAMR)

Require: Source Motion Q^{src}, Robot Models $\mathcal{R}_1, \mathcal{R}_2$
Ensure: Retargeted Motion Q^{rob}
Phase 1: Dual-Manifold Construction
Compute scales: $s_k \leftarrow h_{rob}/h_{src}^{(k)}$, $s_{uni} \leftarrow \text{avg}(s_k)$
for $t = 0 \to T$ **do**
  $p_t^{raw} \leftarrow \text{FK}(q_t^{src})$
  $P_{ind}^{(k)} \leftarrow s_k \cdot p_t^{raw,(k)}$; $P_{uni}^{(k)} \leftarrow s_{uni} \cdot p_t^{raw,(k)}$
  Pre-compute targets: $L_{ref} \leftarrow \mathcal{L}(P_{ind}^{(k)})$
end for
Phase 2: SQP Optimization
Initialize $q_0 \leftarrow q_{nom}$
for $t = 1 \to T$ **do**
  Detect interaction graph $\mathcal{E}_{inter}$ using P_{uni}
  Let $q_{prev} \leftarrow q_{t-1}$
  Formulate QP Subproblem (Solve for Δq):
  Objective: $\min_{\Delta q} \mathcal{J}(q_{prev} + \Delta q)$
    $= w_{self}(\|\mathcal{L}(q) - L_{ref}\|^2 + \lambda_{rot}\|\Delta\theta\|^2)$
    $+ w_{inter} \sum_{(i,j)\in\mathcal{E}} \omega_{ij} \|\Delta p_{ij} - \Delta p_{ij}^{uni}\|^2$
    $+ w_{reg}\|\Delta q\|^2$
  Subject to Constraints:
    1) Limits: $q^{min} \le q_{prev} + \Delta q \le q^{max}$
    2) Collision: $J_{col}\Delta q \ge -\phi(q_{prev}) - \epsilon_{safe}$
    3) Contact: $\|J_{foot}\Delta q\| \le \epsilon_{stick}$ **if** foot_contact
    4) Trust Region: $\|\Delta q\|_2 \le \delta$
  $\Delta q^* \leftarrow \text{OSQP}(\text{Objective}, \text{Constraints})$
  Update: $q_t \leftarrow q_{prev} + \Delta q^*$
  Append q_t to Q^{rob}
end for
return Q^{rob}

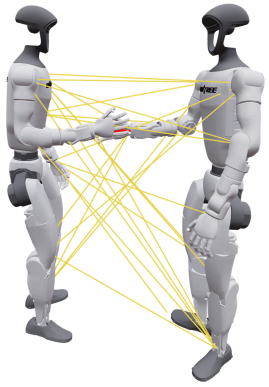

Fig. 7: Visualization of Topological Interaction Priors. We illustrate the extracted graph structures on a representative interaction task, where **yellow edges** denote spatial interaction constraints and **red edges** indicate physical contacts.

By explicitly modeling these connections, IAMR provides topological cues that guide the downstream policy to distinguish between required interaction and unwanted penetration.

B. Details of IGRL

1) Network Architecture: Our policy $\pi_\theta(a_t|o_t)$ employs a hierarchical encoder-decoder architecture designed to process heterogeneous temporal data. The network input is composed of three semantic groups, which are processed by specialized encoders before being fused for action generation.

Observation Space & Inputs
To capture complex coupled dynamics, the policy inputs are organized into two distinct temporal streams: a **Future Window** ($t+1 \dots t+20$) providing feedforward intent, and a **History Window** ($t-19 \dots t$) providing feedback control states.

- **Future Reference ($o_{fut} \in \mathbb{R}^{93}$ per step):** Encodes the look-ahead trajectory for motion planning.
  - *Self Future ($\in \mathbb{R}^{64}$):* Contains the reference joint positions ($\in \mathbb{R}^{29}$) and velocities ($\in \mathbb{R}^{29}$), along with the reference root orientation ($\in \mathbb{R}^{6}$). The orientation is represented by the first two columns of the rotation matrix (Rot6D).
  - *Partner Future ($\in \mathbb{R}^{29}$):* Contains the partner's reference joint positions to anticipate collaborative intent.
- **History Observation ($o_{hist} \in \mathbb{R}^{239}$ per step):** Aggregates the agent's proprioception and perception of the peer.
  - **Proprioception ($o_{prop} \in \mathbb{R}^{157}$):**
    - *Tracking State ($\in \mathbb{R}^{64}$):* Includes the current step's reference joint positions and velocities, combined with the root orientation error ($\in \mathbb{R}^{6}$). The orientation error is computed from the first two columns of the rotation error matrix ($R_{des}^T R_{cur}$).
    - *Physical State ($\in \mathbb{R}^{93}$):* Includes projected gravity ($\in \mathbb{R}^{3}$), base angular velocity ($\in \mathbb{R}^{3}$), joint positions ($\in \mathbb{R}^{29}$), joint velocities ($\in \mathbb{R}^{29}$), and previous actions ($\in \mathbb{R}^{29}$).
  - **Peer Perception ($o_{peer} \in \mathbb{R}^{82}$):**
    - *Partner State ($\in \mathbb{R}^{64}$):* Includes the partner's current reference joint positions ($\in \mathbb{R}^{29}$), actual joint positions ($\in \mathbb{R}^{29}$), and the partner's root orientation error ($\in \mathbb{R}^{6}$).
    - *Interaction Topology ($\in \mathbb{R}^{18}$):* Explicitly encodes the formation geometry by including both the *reference* and *simulation* relative transforms. Each transform consists of the relative position ($p_{rel} \in \mathbb{R}^{3}$) and orientation ($R_{rel} \in \mathbb{R}^{6}$) expressed in the ego-centric frame.

Temporal Encoders (1D-CNN)
We employ two separate 1D-Convolutional Neural Networks to extract features from the observation history and future trajectories. Distinct from standard implementations, we employ an *Input Projection* layer to compress high-dimensional inputs before temporal convolution.

- **Architecture:** Both the History and Future encoders share a dual-layer CNN structure configured as follows:
  - *Input Projection:* Linear mapping to latent dim $C = 60$.
  - *Layer 1:* Conv1d($60 \to 40$, $k=6, s=2$), ELU.
  - *Layer 2:* Conv1d($40 \to 20$, $k=4, s=2$), ELU.
- **Output:** The temporal features are flattened and projected to compact embeddings ($e_{hist} \in \mathbb{R}^{67}, e_{fut} \in \mathbb{R}^{64}$). Uniquely, e_{hist} incorporates an explicit estimation of the base linear velocity ($\hat{v}_{base} \in \mathbb{R}^{3}$) alongside the latent features. This design serves as a fusion of explicit physical estimation and implicit temporal representations, compensating for the lack of direct velocity observations in the real world.

Action Decoder (MLP)
The encoded temporal features are concatenated with the current time-step observation and fed into a Multi-Layer Perceptron (MLP) to generate the action distribution.

- **Structure:** Three hidden layers with $[512, 256, 128]$ units and ELU activation.
- **Output Head:** A linear layer outputs the mean $\mu_t \in \mathbb{R}^{29}$ of the Gaussian distribution for target joint positions. The standard deviation σ is a learnable parameter initialized at 1.0.

2) Reward Definitions: The total reward r_t is computed as a weighted sum of terms designed to balance kinematic fidelity with interaction plausibility, as detailed in Table IV. We prioritize interaction-centric objectives (e.g., relative geometry and contact) over individual tracking precision to encourage compliant multi-agent coupling.

Interaction Graph Reward (r_{inter})
As formulated in the main text (Sec. III-B, *Graph-based Rewards*), this term enforces geometric consistency by penalizing deviations in the interaction edges.

$$r_{inter} = \exp\left(-\frac{1}{\sigma_{inter}} \sum_{(i,j)\in\mathcal{E}} w_{ij} \cdot \| p_{ij}^{sim} - p_{ij}^{ref} \|^2 \right).$$

By inheriting the dynamic weights w_{ij} from the IAMR module, the policy inherently learns to prioritize the same spatial topology (e.g., hand-shoulder proximity) as the optimized reference.

Physical Contact Graph Reward ($r_{contact}$)
Unlike kinematic tracking, physical interaction requires satisfying force constraints that are not present in motion datasets. We propose a graph-based formulation that handles contact nodes dynamically based on their reference status. Leveraging the flexible node definition [42], we map simulation links to a set of abstract contact nodes $\mathcal{V}$ (e.g., palms, feet, pelvis). The reward is computed as a weighted sum of an *active contact* term and an *inactive* term:

$$r_{contact} = \lambda_{act} \cdot e^{-E_{act}/\sigma_c^2} + \lambda_{inact} \cdot e^{-E_{inact}/\sigma_c^2}.$$

The weights λ_{act} and λ_{inact} are adaptive, calculated as the ratio of active/inactive nodes in the current reference frame, ensuring balanced supervision across diverse contact phases.

i) Active Contact Error (E_{act}): For nodes where contact is expected ($k \in \mathcal{V}_{act}$), we define a hybrid error combining binary status consistency and continuous force regularization:

$$E_{act} = \sum_{k\in\mathcal{V}_{act}} \left(\beta \| C_k^{sim} - 1 \| + (1-\beta) \mathcal{L}_{force}(f_k^{sim}) \right).$$

where $C_k^{sim} \in \{0, 1\}$ is the detected contact status. The force regularization term $\mathcal{L}_{force}$ penalizes forces outside the valid range $[F_{min}, F_{max}]$:

$$\mathcal{L}_{force}(f) = \begin{cases} 1.0 - f/F_{min} & \text{if } f < F_{min} \text{ (Too Weak)} \\ (f - F_{max})/F_{max} & \text{if } f > F_{max} \text{ (Too Strong)} \\ 0 & \text{otherwise (Valid)} \end{cases}.$$

This formulation explicitly guides the agent to exert sufficient force for stability ($> F_{min}$) while preventing explosive collisions ($< F_{max}$), solving the ambiguity of "touching without force".

ii) Inactive Contact Error (E_{inact}): For nodes where no contact is expected ($k \notin \mathcal{V}_{act}$), we strictly penalize any detected "ghost interaction":

$$E_{inact} = \sum_{k \notin \mathcal{V}_{act}} \|C_k^{sim} - 0\|.$$

3) Robust Training Strategy: To ensure transferability to the physical world and handle the complexity of coupled interaction phases, we implement a rigorous training protocol comprising error-aware curriculum-based adaptive sampling and extensive domain randomization.

Curriculum-based Adaptive Sampling

Standard Reference State Initialization (RSI) relies on sparse failure counts, which is insufficient for interaction tasks where "survival" does not imply "success" (e.g., maintaining balance but losing interaction geometry). We propose a continuous, error-aware sampling strategy defined as follows:

- **Multi-Objective Error Landscape:** For each discretized motion bin s, we maintain a smoothed error vector $\mathbf{e}(s) = [e_{fail}, e_{track}, e_{inter}]^T$, incorporating failure signals, tracking errors, and interaction violations. To capture temporal causality—where failures stem from preceding suboptimal actions—we smooth these signals using a non-causal Gaussian kernel ($k = 3$) implemented via 1D convolution.
- **Dynamic Probability:** The sampling probability $P(s)$ is a convex combination of uniform exploration ($\eta = 0.05$) and error-weighted exploitation, modulated by curriculum weights $\boldsymbol{\alpha}$:

$$P(s) = \eta \frac{1}{S} + (1 - \eta) \sum_k \alpha_k(\bar{L}_{max}) \frac{e_k(s)}{\sum_j e_k(j)}.$$

- **Curriculum Schedule:** The weight vector $\boldsymbol{\alpha}$ evolves based on the maximum moving average episode length $\bar{L}_{max}$ to ensure monotonic progress:
  - *Stability Phase ($\bar{L}_{max} < 350$):* Weights are fixed at $\boldsymbol{\alpha}_{init} = [0.8, 0.1, 0.1]$. The policy prioritizes states leading to terminal failures to learn basic balance capabilities.
  - *Transition Phase ($350 \leq \bar{L}_{max} < 500$):* We perform **linear interpolation** between $\boldsymbol{\alpha}_{init}$ and the target weights $\boldsymbol{\alpha}_{target} = [0.05, 0.30, 0.65]$. This gradually shifts the focus from avoiding failures to minimizing tracking and interaction errors as the agent gains proficiency.
  - *Precision Phase ($\bar{L}_{max} \geq 500$):* Weights stabilize at $\boldsymbol{\alpha}_{target}$. The sampling strictly targets complex interaction phases that impose high control complexity despite low failure probability.

Dual-Agent Domain Randomization

To bridge the Sim-to-Real gap, we introduce perturbations targeting both physical dynamics and the specific challenges of distributed multi-agent communication, as detailed in Table V.

Communication & Initialization Strategy. Beyond standard dynamics randomization, we implement specific strategies for dual-agent coordination:

- **Communication Degradation:** We explicitly model the latency in distributed systems. By training with randomized delays ($20 \sim 60$ms) for both peer proprioception (wireless transmission) and relocalization (vision processing), the policy learns to be robust against asynchronous data reception.
- **Initial State Perturbation:** To handle calibration errors, we initialize episodes with random offsets in root position ($\pm$5cm) and orientation ($\pm$0.2rad). This forces the policy to learn active recovery behaviors from suboptimal relative configurations immediately upon activation.

C. Experimental Setup & Metrics

1) Baseline Implementation Details: To strictly validate our contributions, we benchmark our framework against two sets of baselines: kinematic retargeting methods (answering Q1) and dynamic policy learning variants (answering Q2).

Retargeting Baselines (Kinematic Level)

All baselines utilize the same source motion data and undergo identical skeletal scaling pre-processing to ensure fair comparison.

- **GMR (General Motion Retargeting) [3]:** A standard Cartesian-space optimization method. It treats the two humanoids as isolated entities, optimizing joint angles to minimize the tracking error of individual keypoints' **positions and rotations** (e.g., end-effectors, pelvis) relative to the source motion. It does not include any interaction-aware constraints.
- **OR (OmniRetarget) [49]:** A retargeting method utilizing Interaction Mesh (I-Mesh) to preserve geometric topology, originally designed for Human-Object Interaction (HOI) scenarios. In our dual-agent setting, we apply OR to each agent independently. While it preserves individual body topology, it lacks mechanisms to model or preserve inter-agent spatial relationships.
- **DOR (Dual-OmniRetarget):** A strong baseline we constructed by extending OmniRetarget to the multi-agent setting. We construct a holistic interaction mesh that encompasses both robot bodies, creating "cross-agent" edges between proximal body parts. However, by treating the dual-agent system as a single unified mesh, it fails to distinguish between critical inter-agent interactions and

TABLE IV: Reward Terms and Weights used in IGRL

Term	Weight	Equation	Description
Interaction Graph Objectives			
Interact Edge	1.5	$\exp\left(-\frac{1}{\sigma_i}\sum_{(i,j)\in\mathcal{E}} w_{ij}\|p_{ij}^{sim} - p_{ij}^{ref}\|^2\right)$	Enforces geometric consistency of interaction edges.
Contact	1.0	$\lambda_{act}e^{-E_{act}/\sigma_c^2} + \lambda_{inact}e^{-E_{inact}/\sigma_c^2}$ **where** $E_{act} = \sum_{k\in\mathcal{V}_{act}}(\beta\|C_k^{sim} - 1\| + (1-\beta)\mathcal{L}_f)$ **and** $E_{inact} = \sum_{k\notin\mathcal{V}_{act}}\|C_k^{sim} - 0\|$	Balances active contact enforcement (force constrained) and ghost interaction suppression.
Motion Tracking Objectives			
Upper Pos	1.0	$\exp(-\frac{1}{N_u}\sum_{k\in Upper}\|p_k^{sim} - p_k^{ref}\|^2/\sigma_{pos}^2)$	Tracks Euclidean positions of upper body links.
Upper Ori	1.0	$\exp(-\frac{1}{N_u}\sum_{k\in Upper}\|\log((R_k^{sim})^\top R_k^{ref})\|^2/\sigma_{ori}^2)$	Tracks orientation of upper body links.
Upper Lin Vel	1.0	$\exp(-\frac{1}{N_u}\sum_{k\in Upper}\|v_k^{sim} - v_k^{ref}\|^2/\sigma_{vel}^2)$	Matches linear velocities of upper body.
Upper Ang Vel	1.0	$\exp(-\frac{1}{N_u}\sum_{k\in Upper}\|\omega_k^{sim} - \omega_k^{ref}\|^2/\sigma_{ang}^2)$	Matches angular velocities of upper body.
Lower Pos	0.5	$\exp(-\frac{1}{N_l}\sum_{k\in Lower}\|p_k^{sim} - p_k^{ref}\|^2/\sigma_{pos}^2)$	Tracks Euclidean positions of lower body links.
Lower Ori	0.5	$\exp(-\frac{1}{N_l}\sum_{k\in Lower}\|\log((R_k^{sim})^\top R_k^{ref})\|^2/\sigma_{ori}^2)$	Tracks orientation of lower body links.
Lower Lin Vel	0.5	$\exp(-\frac{1}{N_l}\sum_{k\in Lower}\|v_k^{sim} - v_k^{ref}\|^2/\sigma_{vel}^2)$	Matches linear velocities of lower body.
Lower Ang Vel	0.5	$\exp(-\frac{1}{N_l}\sum_{k\in Lower}\|\omega_k^{sim} - \omega_k^{ref}\|^2/\sigma_{ang}^2)$	Matches angular velocities of lower body.
Anchor Pos	0.3	$\exp(-\|p_{root}^{sim} - p_{root}^{ref}\|^2/\sigma_{root}^2)$	Tracks root position in world frame to prevent global drift.
Anchor Ori	0.5	$\exp(-\|\log((R_{root}^{sim})^\top R_{root}^{ref})\|^2/\sigma_{root\_ori}^2)$	Tracks root heading in world frame.
Regularization & Penalties			
Action Rate	-0.3	$\|\mathbf{a}_t - \mathbf{a}_{t-1}\|^2$	Penalizes action changes to ensure smooth control.
Feet Slip	-0.5	$\sum_{k\in Feet}\mathbb{I}(contact_k)\cdot\|v_{k,xy}\|^2$	Penalizes foot sliding velocity during ground contact.
Joint Limit	-10.0	$\sum_j \max(0, \|q_j\| - q_{limit})$	Penalizes violations of physical joint limits.
Torque	10^{-4}	$-\|\tau\|^2$	Prevents excessive torques.

TABLE V: Domain Randomization Parameters

Term	Value
Dynamics Randomization	
Link Mass	$\mathcal{U}[0.9, 1.1]\times$ default (per link)
CoM Offset (Torso)	$\Delta x, y, z \in \mathcal{U}[-0.05, 0.05]$ m
Joint Friction	Static: $\mathcal{U}[0.3, 2.0]$, Dynamic: $\mathcal{U}[0.3, 1.6]$
Actuator Gains	Stiffness/Damping: $\mathcal{U}[0.9, 1.1]$
Restitution	$\mathcal{U}[0.0, 0.8]$ (Ground contact)
Default Joint Pos	$\Delta\theta_0 \sim \mathcal{U}[-0.01, 0.01]$ rad (Calibration)
Control Delay	$\mathcal{U}[0, 15]$ ms
External Perturbations	
Robot Push (Linear)	$v_{x,y} \sim \mathcal{U}[-0.4, 0.4]$ m/s, $v_z \sim \mathcal{U}[-0.16, 0.16]$ m/s
Robot Push (Angular)	$\omega_{x,y} \sim \mathcal{U}[-0.4, 0.4]$ rad/s, $\omega_z \sim \mathcal{U}[-0.64, 0.64]$ rad/s
Push Interval	Applied every $1.0 \sim 3.0$ s
Initial Pose Offset	$\Delta pos \in [-5, 5]$ cm, $\Delta yaw \in [-0.2, 0.2]$ rad
Communication Degradation	
Peer Latency	Proprioception & Relocalization: $\mathcal{U}[20, 60]$ ms

individual self-motion, leading to an optimization that lacks focus on core interaction information.

Policy Learning Baselines (Dynamic Level)

To isolate the contribution of specific components in our IGRL framework, we compare against the following variants. All variants are trained using the same PPO hyperparameters and network architecture unless specified otherwise.

- **Single Agent:** Represents the "Status Quo" approach. This policy treats the peer robot merely as a dynamic obstacle or ignores it entirely.
  Implementation: The peer observation o_{peer} is masked, and the reward function includes only standard tracking terms (r_{track}) without any interaction (r_{inter}) or contact ($r_{contact}$) objectives. The agent is rewarded solely for tracking its own retargeted reference.
- **Ours w/o Peer Obs:** Evaluates the necessity of explicit inter-agent perception.
  Implementation: The policy architecture is identical to IGRL, but the peer observation stream o_{peer} is zeroed out. The reward function remains full (including interaction rewards), forcing the agent to infer interaction requirements solely from its own proprioception and the reference motion.
- **Ours w/o Contact Rew:** Evaluates the contribution of the Physical Contact Graph.
  Implementation: The policy is trained with the full observation space, but the contact reward $r_{contact}$ weight is set to zero. This tests whether kinematic tracking alone is sufficient to establish stable physical coupling.
- **Ours w/o Interaction Rew:** Evaluates the contribution of the Spatial Interaction Graph.
  Implementation: The interaction graph reward r_{inter}

is removed. The agent relies purely on standard position/rotation tracking rewards to maintain formation. This tests the importance of explicitly optimizing relative topology for resolving geometric ambiguities.

2) Evaluation Metric Implementation Details: We employ specific metrics for each component of our framework to evaluate kinematic quality and dynamic performance respectively.

Retargeting Evaluation Metrics (Q1)
We evaluate retargeting quality from three complementary aspects: physical feasibility, interaction fidelity, and downstream utility.

- **Inter-Penetration Rate (IPR) & Max Penetration Depth (MPD):** Measure physical feasibility by reporting the percentage of frames exhibiting inter-agent penetration and the maximum penetration depth across the sequence. Lower values indicate safer motion.
- **Interaction Edge Error (IEE):** Quantifies geometric interaction fidelity as the normalized L2 distance between retargeted interaction edges and the scaled ground-truth edges.
- **Contact F1 Score:** Evaluates binary contact accuracy against ground-truth interaction edges following [20, 29]. We report F1 scores under **Strict** ($\tau < 0.2\,\text{m}$) and **Loose** ($\tau < 0.4\,\text{m}$) contact settings to assess precision at different scales.
- **Downstream Success Rate (DSR):** Serves as a proxy for physical learnability. It is defined as the percentage of RL rollouts where the agent can track the retargeted reference while maintaining the interaction structure (i.e., maintaining an IEE deviation $< 20\%$ relative to the scaled ground truth).

Policy Learning Evaluation Metrics (Q2)
To assess the robustness and fidelity of the learned control policy in dynamic environments, we evaluate interaction and contact performance with respect to the retargeted reference trajectories.

- **Interaction Performance (ISR & IEE):** Measures interaction fidelity by computing the distance-weighted relative error between simulated interaction edges and the retargeted reference edges. We report the mean **Interaction Edge Error (IEE)** and the **Interaction Success Rate (ISR)**, defined as the percentage of steps where the IEE is kept within a strict tolerance ($< 10\%$).
- **Contact Performance (CSR & CER):** Evaluates physical contact fidelity relative to the retargeted reference contacts. The **Contact Error Rate (CER)** measures the rate of violations of required contact constraints, while the **Contact Success Rate (CSR)** reports the percentage of steps where sufficient reference contacts ($> 80\%$) are correctly recalled by the policy.

3) Sim-to-Real Hardware: We validate our approach on the Unitree G1 humanoid robot platform. To bridge the gap between ideal simulation states and real-world noisy sensor data, we implement a fully onboard perception and control stack written in C++ for real-time performance.

Robot Platform & Compute. The Unitree G1 (approx. 1.3 m height, 29 DoF) serves as our experimental testbed. Unlike simulation where state information is privileged, all computations—including state estimation, policy inference, and low-level control—are performed onboard the robot's internal CPU. We utilize ONNX Runtime to execute the trained policies, achieving an inference latency of less than 3 ms, ensuring a stable 50 Hz control loop.

Hierarchical Localization System. Precise relative localization is a prerequisite for our *Interaction Graph* reward calculation. We develop a robust, coarse-to-fine localization framework fusing LiDAR and IMU data:

- **High-Frequency Odometry:** We utilize **Point-LIO** [14], a robust LiDAR-Inertial Odometry framework, to provide high-bandwidth (10 Hz) state estimation robust to aggressive motions and vibrations.
- **Global Initialization (Coarse):** To handle the "cold start" problem and global drift, we implement a neural registration service based on **GeoTransformer**. This module extracts superpoint features to perform global registration between the current scan and a pre-built point cloud map, providing a reliable initial pose guess.
- **Real-Time Tracking (Fine):** During operation, a C++ relocalization node refines the pose using **GICP** (Generalized Iterative Closest Point) [35]. A Kalman Filter fuses the high-frequency relative odometry from Point-LIO with the low-frequency global pose corrections from GICP to output smooth, drift-free global states.

Control Architecture. The deployment system operates on a dual-frequency hierarchy to match the simulation setup:

- **High-Level Policy (50 Hz):** The motion tracking policy constructs observations based on the fused state estimates and computes desired joint positions.
- **Low-Level Controller (500 Hz):** A real-time PD controller converts these targets into motor torques, enforcing the physical compliance required for safe interaction.

Multi-Agent Synchronization. To enable coordinated interaction, the agents exchange their estimated global poses and full peer observations (o_{peer}) via the **LCM** (Lightweight Communications and Marshalling) protocol [21].