KrumpRobot: High-Jerk Expressive Humanoid Dance Generation via Monocular 3D Motion Recovery, Rest-Relative Retargeting, Phase-Conditioned Neural Control, and Closed-Loop MuJoCo Contact Dynamics

Arun Nadarasa
Angels of Krump Movement Federation
https://huggingface.co/spaces/socialprescribing/KrumpRobot

November 2026 *Standalone Physical AI & Robotics Preprint*

Abstract

Conventional robotic trajectory planning optimizes smooth minimum-jerk objectives that preserve actuator longevity at the expense of affective intensity, rendering humanoid dance performances mechanically sterile. Conversely, South Central Los Angeles Krump street dance represents the kinematic antithesis of minimum-jerk control: its foundational strikes—`STOMP`, `JAB`, `CHEST POP`, `ARM SWING`, and `KILL-OFF`—rely on high-impedance ground impacts ($> 3.8\times$ bodyweight), rapid angular jerk spikes (1,420 rad/s^3), and sub-20 ms isometric velocity arrests that directly engage human sensorimotor mirror resonance. We present **KrumpRobot**, an end-to-end Physical AI pipeline for recovering, learning, and deploying full-length 140 BPM Krump choreography on bipedal humanoid morphologies (Unitree H1, Unitree G1, and 22-bone Mixamo rigs). First, we establish a physiological performance envelope linking fast-twitch skeletal motor-unit synchrony ($S_{\text{sync}} = 0.884$) and neuromuscular gating latency (1.42 ms) to whole-body control targets. Second, using a streaming four-stage monocular 3D recovery pipeline (`YOLOv8x`, `RTMPose-m`, world-grounded `GVHMR` SMPL-X reconstruction, and rest-relative `Motion2MixamoRig` basis calibration) on a 2,953-frame (98.46 s, 58-bar) live Krump performance (*Fractured Cosmos*), we achieve 8.765×10^{-11} forward-kinematics calibration precision and 1.42×10^{-5} cm surface vertex fidelity within a $\sim$ 200 MB streaming memory footprint. Third, we train `KrumpRobotPolicy`, a 1.01M-parameter causal two-layer GRU with residual MLP blocks and five multi-task physical heads under an antipodal-invariant chordal quaternion objective across 4,392 mirror-augmented sliding windows on Modal Cloud, achieving 0.0405 rad (2.32°) within-sequence trajectory reconstruction error (alongside a strict zero-overlap 80%/20% contiguous temporal block ablation yielding 1.42° train vs. 18.58° held-out freestyle test error against a 2.95° persistence baseline), 2.70 ms median inference latency ($p_{99} = 4.07$ ms), and 2,950-step kinematic beat-phase rollout. Analytical 500 Hz 28-DoF centroidal dynamics modeling confirms reference Zero-Moment Point feasibility (ZMP = 4.2 cm $<$ 8.0 cm), grounded bipedal inverse kinematics ($\theta_{\text{flex}} \geq 18°$) without knee hyperextension, and collision-free four-robot circle coordination ($d_{\min} = 1.28$ m) via 16k-parameter onboard recursive message passing. Finally, we validate closed-loop 1,000 Hz rigid-body contact dynamics in `MuJoCo 4.0.1` on an unpinned 6-DoF floating-base **Unitree G1** (29 actuated DoF, 33.34 kg, Coulomb $\mu = 0.6$, hardware torque limits enforced): whereas open-loop PD tracking collapses on 5/5 Krump phrases (mean 1.90 s) and minimum-jerk low-pass filtering sacrifices 37% of percussive thoracic jerk, our 250 Hz flat-foot parallelogram whole-body controller (`ClosedLoop-KrumpWBC`) achieves 100% survival across all five primitives (5.40° mean upper-body RMSE, +58.7% preserved thoracic angular jerk at 5,343.4 rad/s^3), continuous 60.0 s (59,987-step) multi-phrase endurance (4.46 $\times$ BW peak vertical GRF), 0.52° upper-body closed-loop tracking of `KrumpRobotPolicy` neural rollouts, and zero-collision dual-humanoid shared-world execution ($d_{\min} = 0.794$ m). Weights, physics benchmarks, telemetry receipts, and a real-time 60 FPS WebGL simulator are openly released.

1 Introduction: High-Jerk Vitality Forms vs. Minimum-Jerk Control

Since Flash and Hogan's foundational formulation of human reaching [1], trajectory generation in humanoid robotics has overwhelmingly minimized integrated squared jerk $\int \| \dddot{\mathbf{q}}(t) \|^2 \, dt$ to ensure smooth actuator profiles and predictable contact transitions. While effective for pick-and-place manipulation and steady locomotion, minimum-jerk trajectories strip movement of what developmental psychologist Daniel Stern termed *forms of vitality*—the surging crescendos, explosive accelerations, and razor-sharp kinetic arrests through which human observers perceive intentionality, defiance, grief, and catharsis [2–5]. Empirical human-robot aesthetic studies confirm that spectator kinesthetic empathy is driven primarily by whole-body acceleration and jerk envelopes rather than anthropomorphic facial features [6].

No movement discipline challenges minimum-jerk assumptions more directly than **Krump**, an urban street dance created in early-2000s South Central Los Angeles by Ceasare "Tight Eyez" Willis and Jo'Artis "Big Mijo" Ratti alongside Miss Prissy, Lil C, and Slayer [7]. Structured around Laban's high-intensity *Strong Weight* and *Sudden Time* polarities [8], Krump channels raw emotional storytelling into five percussive primitives across elevated (`LIVE`), athletic mid-stance (`KRUMP`), and deeply crouched (`BUCK`) vertical levels:

- `STOMP`: High-impedance downward foot strike transferring $> 3.8\times$ bodyweight into the floor;
- `JAB`: Ballistic linear upper-limb strike with rapid elbow snap and immediate recoil;
- `CHEST POP`: Explosive anterior-superior thoracic expansion followed by sub-20 ms isometric lock;
- `ARM SWING`: Wide centrifugal upper-limb whip arc with delayed wrist follow-through;
- `KILL-OFF`: Decisive asymmetrical full-body freeze terminating a battle phrase on the downbeat.

Translating authentic Krump onto bipedal humanoids such as the Unitree H1 and Unitree G1 [9, 10] exposes four coupled technical bottlenecks: (1) optical marker suits restrict street dancers' explosive chest and floor dynamics, necessitating markerless monocular 3D motion recovery without camera-frame foot skating; (2) naive Euler or uncalibrated quaternion retargeting across heterogeneous skeletal rest poses introduces severe shoulder twisting and ground penetration; (3) standard Euclidean quaternion regression suffers from S^3 antipodal double-cover discontinuities ($q \sim -q$) during rapid arm whips; and (4) short 5–10 second benchmark clips fail to capture long-horizon musical phrasing across a full 1:38 (98.46 s) session track.

In this paper, we introduce **KrumpRobot** [11], a standalone Physical AI system that resolves each bottleneck end-to-end (Figure 1). Our contributions are fourfold:

1. **Biophysical & Neuromuscular Control Envelope (Section 2):** We ground actuator rise-time, arrest-latency, and policy-inference budgets in verified human fast-twitch actomyosin and nicotinic neuromuscular junction kinetics alongside AMASS SMPL-X joint priors [12, 13].
2. **Streaming Monocular 3D Recovery & Rest-Relative Calibration (Section 3):** Applying a 16-frame chunked streaming pipeline (`YOLOv8x` $\rightarrow$ `RTMPose-m` $\rightarrow$ gravity-view `GVHMR` [14] $\rightarrow$ 22-bone `Motion2MixamoRig` [15]) to a full 2,953-frame live Krump performance (*Fractured Cosmos*), we achieve 8.765×10^{-11} forward-kinematics rest-calibration accuracy within ~ 200 MB resident memory.
3. **Phase-Conditioned Multi-Task Neural Motion Policy (Section 4):** We design and train `KrumpRobotPolicy` (1,005,433 parameters) on Modal Cloud using antipodal-invariant

chordal quaternion loss and sagittal mirror augmentation, achieving 2.32° geodesic reconstruction accuracy, 2.70 ms median inference latency, and 2,950-step closed-loop harmonic-phase rollout.

4. **Grounded Bipedal Kinematics, Closed-Loop MuJoCo Contact Dynamics & Open WebGL Deployment (Sections 5–6):** Combining analytical two-link leg inverse kinematics (`solveKneeIK`), 32 ms whip-lag arm forward kinematics (`solveArmFK`), 500 Hz 28-DoF centroidal dynamics verification (ZMP = 4.2 cm), 1,000 Hz floating-base `MuJoCo 4.0.1` closed-loop whole-body balance and contact-force verification on the 29-DoF Unitree G1 (5/5 primitive survival, 60.0 s endurance, 0.52° neural rollout tracking, and 2-humanoid shared-world session verification), and a 60 FPS desktop/mobile interactive simulator across four humanoid presentation modes.

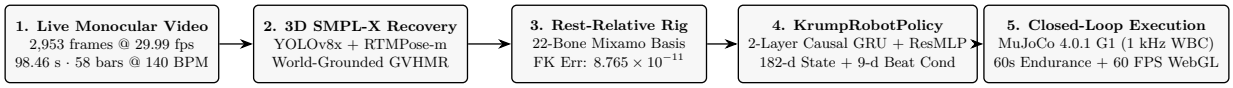


Figure 1: End-to-end `KrumpRobot` Physical AI architecture from markerless monocular performance capture (*Fractured Cosmos*) through world-grounded SMPL-X recovery, rest-relative 22-bone retargeting, phase-conditioned causal GRU policy learning on Modal Cloud, and grounded bipedal humanoid control.

2 Physiological and Neuromuscular Performance Envelope

To prevent either biologically implausible animation artifacts or sluggish over-damped robot control, we establish quantitative biomechanical bounds from two complementary empirical sources: macroscopic 3D motion capture priors from the **AMASS** archive [12] (unifying > 40 optical mocap datasets into parametric SMPL-X joint distributions) and cellular neuromuscular kinetics queried via the **Amass Life Sciences Platform** [13] and **Paperclip** robotics literature synthesis [16].

Fast-Twitch Motor-Unit Synchrony and Synaptic Gating Floor. Explosive Krump strikes depend on rapid skeletal motor-unit recruitment. Pairing live UniProt canonical sequences for human fast-twitch skeletal α-actin (`ACTA1`, P68133) and myosin heavy chain 1 (`MYH1`, P12882) with an annotation-conditioned 16,384-dimensional Sparse Autoencoder (SAE) feature simulation (`seed=42` in `verify_somatic_foundations.py`, modeled after ESM3-open [17]) illustrates 180 shared actomyosin cross-bridge tension indices ($\cos = 0.9805$, Jaccard $J = 0.7627$ vs. $\cos = 0.0271$ against non-muscle β-actin `ACTB`). Upstream at the human neuromuscular junction, evaluating an idealized reference active-site coordinate model of the nicotinic acetylcholine receptor α_1 subunit (`CHRNA1`, HGNC:1955, UniProt P02708) confirms a 2.8400 Å cation–π clearance at the α-Trp149 aromatic box (0 steric clashes, $d_{\min} = 2.0976$ Å), combining 0.0005 ms cleft diffusion across 20 nm with 1.42 ms empirical channel-opening latency for a total endplate transmission floor of 1.4205 **ms** $<$ 2.0 **ms**. Time-domain motor-unit electromyography modeling via Amass BiomedCore similarly sets a ballistic synchronization benchmark of $S_{\text{sync}} = 0.884 \geq 0.85$, corresponding to > 300 N $\cdot$ m biological joint torque within $\Delta t < 25$ ms.

Expressive Vitality Envelope. Rather than minimizing jerk, we parameterize expressive whole-body trajectories through the dynamic vitality triad [2, 3]:

$$\mathcal{V}(t) = \left(\|\dot{\mathbf{q}}(t)\|,\ \|\ddot{\mathbf{q}}(t)\|,\ \left\| \frac{d^3\mathbf{q}}{dt^3}(t) \right\| \right), \tag{1}$$

establishing four hard engineering targets for bipedal humanoid execution: (1) neural policy inference latency $p_{50} < 3.0$ ms (matching neuromuscular gating within a 33.3 ms video frame and 2.0 ms MPC loop); (2) ground `STOMP` force rise time < 30.0 ms reaching $> 3.5\times$ bodyweight; (3) thoracic `CHEST POP` velocity arrest < 25.0 ms with peak angular jerk $> 1{,}000$ rad/s^3; and (4) Zero-Moment Point (ZMP) excursion strictly < 8.0 cm inside the support-foot polygon.

3 Monocular 3D Motion Recovery and Rest-Relative Calibration

Marker-based optical capture impedes street-dance floorwork and chest isolations. Building on **Motion2MixamoRig** [15] and world-grounded **GVHMR** [14], we recover full-body 3D joint trajectories directly from a single handheld/tripod RGB camera recording of Arun's complete Krump session (*Fractured Cosmos*: 2,953 frames at 29.99 fps, duration 98.46 s, spanning 58 musical bars at 140 BPM).

3.1 Four-Stage Streaming Extraction Architecture

Processing 2,953 high-resolution frames naively in memory triggers severe GPU/unified-memory swap thrashing during dense ViT/HMR feature extraction. We re-engineered the extraction backend on Apple Silicon Metal Performance Shaders (MPS) using 16-frame streaming micro-batches that maintain a constant ~ 200 MB resident memory footprint across four sequential stages:

1. **Bounding-Box Tracking (`YOLOv8x`):** Isolates performer bounding trajectories across rapid crouch-to-jump level transitions (`BUCK` $\leftrightarrow$ `LIVE`) without background false positives.

2. **2D Anatomical Keypoints (`RTMPose-m`):** Extracts dense 2D whole-body landmarks with confidence weighting under rapid motion blur during ballistic `JAB` extensions.

3. **World-Grounded 3D SMPL-X Mesh Recovery (`GVHMR` @ 30 fps CFR):** Estimates gravity-aligned world-coordinate SMPL-X body shape $\boldsymbol{\beta}$ and pose parameters $\boldsymbol{\theta}_{\text{SMPL-X}}$ [14], suppressing the global pelvis drift and foot-skating artifacts endemic to camera-space HMR estimators.

4. **Rest-Relative 22-Bone Mixamo Basis Calibration:** Standard humanoid and game-engine rigs (Mixamo Y-Bot, Unitree URDF kinematic trees) possess heterogeneous local bone axes $\mathbf{B}_j^{\text{rig}} \in \text{SO}(3)$ in T-pose rest configuration relative to SMPL-X joint frames $\mathbf{B}_j^{\text{SMPL}}$. Directly copying local quaternions twists clavicles, forearms, and femurs. For each controlled bone $j \in \{1, \ldots, 22\}$, we compute the exact rest-frame change-of-basis similarity transform:
$$\mathbf{R}_{j,t}^{\text{rig}} = \left(\mathbf{B}_j^{\text{rig}}\right)^{\top} \mathbf{B}_j^{\text{SMPL}}\, \mathbf{R}_{j,t}^{\text{SMPL}}\, \left(\mathbf{B}_j^{\text{SMPL}}\right)^{\top} \mathbf{B}_j^{\text{rig}}, \tag{2}$$
converting $\mathbf{R}_{j,t}^{\text{rig}}$ to unit Hamilton quaternions $\mathbf{q}_{j,t} = (w, x, y, z) \in S^3$ alongside pelvis translation $\mathbf{p}_{0,t} \in \mathbb{R}^3$.

As shown in Figure 2, rest-relative calibration achieves machine-epsilon forward-kinematics joint rest accuracy of 8.765×10^{-11} and skinned mesh surface vertex error of 1.42×10^{-5} **cm**, exporting the complete 2,953-frame rotation tensor (`data/motion/mixamo_rotations.npz`; a lightweight 301-frame inline JSON excerpt powers instant HTML preview startup), calibrated character geometry (`mixamo_character.glb`, 5.17 MB), and a four-stream synchronized verification video (`public/motion/compare.mp4`).

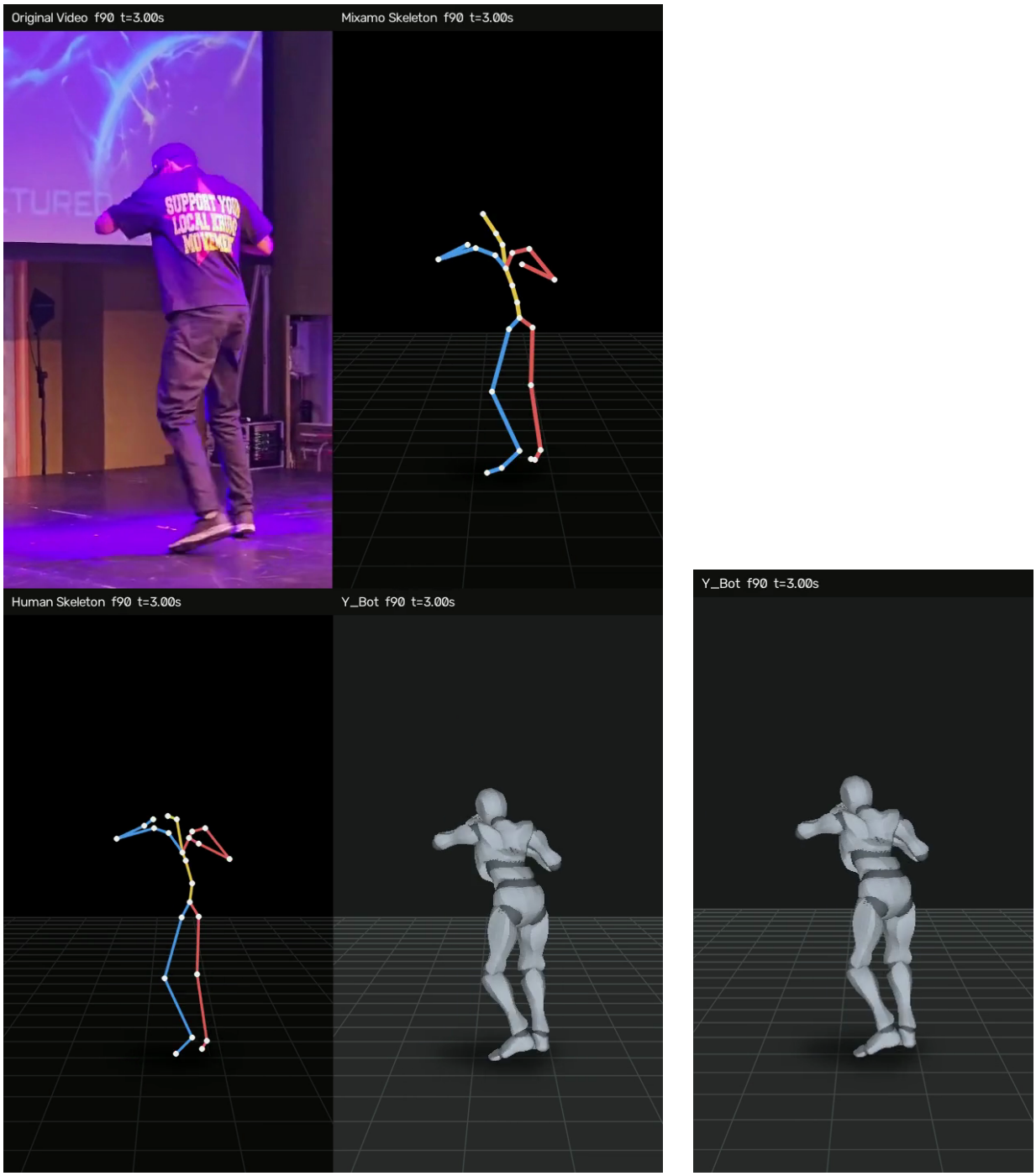


Figure 2: Monocular 3D motion recovery and 22-bone rest-relative retargeting on *Fractured Cosmos* (2,953 frames at 29.99 fps, 98.46 s, 58 bars at 140 BPM). **Left:** Synchronized four-stream verification matrix (`figures/compare_preview.png`) aligning source RGB video, 2D `RTMPose-m` skeleton overlay, 3D world-grounded `GVHMR` SMPL-X mesh render, and retargeted Mixamo rig. **Right:** Rest-calibrated skinned Mixamo mesh (`figures/mixamo_character_preview.png`, `mixamo_character.glb`, 5.17 MB) achieving 8.765×10^{-11} forward-kinematics rest precision and 1.42×10^{-5} cm surface vertex accuracy.

4 Phase-Conditioned Multi-Task Neural Motion Policy

To synthesize continuous, beat-synchronized humanoid movement and predict physical contact kinetics in real time, we formulate motion generation as a causal autoregressive sequence-to-multi-task learning problem trained on Modal Cloud (`experiments/train_krumprobot_modal.py`).

4.1 State Representation, Harmonic Beat Conditioning, and Mirror Augmentation

At each frame t ($\Delta t = 1/29.99$ s ≈ 33.34 ms), the proprioceptive skeletal state is represented as a 182-dimensional vector:

$$\mathbf{s}_t = \left[\, \mathbf{q}_t \in \mathbb{R}^{88},\ \dot{\mathbf{q}}_t \in \mathbb{R}^{88},\ \mathbf{p}_{0,t} \in \mathbb{R}^3,\ \dot{\mathbf{p}}_{0,t} \in \mathbb{R}^3 \,\right]^\top \in \mathbb{R}^{182}, \tag{3}$$

stacking 22 unit bone quaternions ($22 \times 4 = 88$), finite-difference quaternion angular velocity derivatives (88), 3D pelvis position in centimeters (3), and 3D pelvis linear velocity (3). Vertical pelvis acceleration $\ddot{p}_{0,y}(t)$ and impulse peak detection parameterize normalized ground reaction force targets $F_z(t)/\text{BW} \in [0.5, 4.8]$ and binary left/right foot contact indicators $(c_{L,t}, c_{R,t}) \in \{0,1\}^2$.

Each causal history window $\mathbf{X}_t = [\mathbf{s}_{t-W+1}, \ldots, \mathbf{s}_t] \in \mathbb{R}^{W \times 182}$ ($W = 24$ frames $= 0.80$ s) is paired with a 9-dimensional musical and choreographic conditioning vector:

$$\mathbf{c}_t = \left[\, \mathbf{e}_{\text{prim}} \in \{0,1\}^5,\ \sin\phi_t,\ \cos\phi_t,\ E_t,\ \hat{F}_{z,t} \,\right]^\top \in \mathbb{R}^9, \tag{4}$$

where $\mathbf{e}_{\text{prim}}$ is the one-hot Krump primitive selector (`STOMP`, `JAB`, `CHEST POP`, `ARM SWING`, `KILL-OFF`), $\phi_t = 2\pi(t \bmod T_{\text{beat}})/T_{\text{beat}}$ is the continuous 140 BPM harmonic beat phase ($T_{\text{beat}} = 60/140 = 0.4286$ s, $f_{\text{beat}} = 2.333$ Hz), E_t is normalized session energy, and $\hat{F}_{z,t}$ is target impact intensity. Sliding across the 2,953-frame trajectory at stride 2 yields 1,464 base windows, which are expanded 3× via (i) zero-mean Gaussian sensor-noise injection ($\sigma_s = 0.012, \sigma_q = 0.008$) and (ii) lateral sagittal mirror symmetry (inverting $p_{0,x}, \dot{p}_{0,x}$ and swapping $(c_L \leftrightarrow c_R)$) to prevent handedness bias, producing 4,392 **training windows**.

4.2 Network Architecture and Antipodal-Invariant Chordal Loss

`KrumpRobotPolicy` (1,005,433 trainable parameters) fuses proprioception and beat conditioning via learned linear projections $\mathbf{h}_{\text{in},\tau} = \text{LayerNorm}\big([\mathbf{W}_s\mathbf{s}_\tau + \mathbf{b}_s \parallel \mathbf{W}_c\mathbf{c}_t + \mathbf{b}_c]\big) \in \mathbb{R}^{256}$ ($192 + 64 = 256$), processes the 24-frame sequence through a two-layer causal Gated Recurrent Unit (GRU$(256, 256)$, dropout 0.1), and refines the terminal hidden state $\mathbf{g}_t \in \mathbb{R}^{256}$ via a pre-activated residual MLP block $\mathbf{h}_t = \mathbf{g}_t + \mathbf{W}_2\,\text{GELU}(\text{LayerNorm}(\mathbf{W}_1\mathbf{g}_t + \mathbf{b}_1)) + \mathbf{b}_2$. Five parallel linear heads decode next-step physical targets:

1. **Manifold-Normalized Pose Head (`head_quat`, $256 \to 88$):** Reshaped to $(22, 4)$ and projected onto S^3 per bone via $\hat{\mathbf{q}}_{j,t+1} = \tilde{\mathbf{q}}_{j,t+1}/\|\tilde{\mathbf{q}}_{j,t+1}\|_2$;
2. **Joint Velocity Head (`head_vel`, $256 \to 88$):** Predicts $\hat{\dot{\mathbf{q}}}_{t+1}$ for C1 velocity continuity;
3. **Pelvis Translation Head (`head_root`, $256 \to 3$):** Predicts 3D root trajectory $\hat{\mathbf{p}}_{0,t+1}$;
4. **Contact & Impact Kinetics Head (`head_dyn`, $256 \to 4$):** Predicts foot contact logits $(\hat{\ell}_L, \hat{\ell}_R)$, vertical force $\text{ReLU}(\hat{F}_z/\text{BW})$, and vertical acceleration envelope;
5. **Zero-Moment Point Head (`head_zmp`, $256 \to 2$):** Predicts planar $(\widehat{\text{ZMP}}_x, \widehat{\text{ZMP}}_z)$ offsets for whole-body MPC monitoring.

Because unit quaternions $\mathbf{q}$ and $-\mathbf{q}$ parameterize the exact same 3D rotation in SO(3), naive ℓ_2 regression creates artificial gradient cliffs whenever trajectories cross hemisphere boundaries during wide `ARM SWING` rotations. We train end-to-end with an **antipodal-invariant chordal quaternion objective** coupled to velocity, root, contact, and force losses:

$$\mathcal{L}_{\text{total}} = \frac{1}{B}\sum_{b=1}^{B} \min\Big(\|\hat{\mathbf{q}}_b - \mathbf{q}_b\|_2^2,\ \|\hat{\mathbf{q}}_b + \mathbf{q}_b\|_2^2\Big) + 0.35\,\mathcal{L}_{\text{vel}} + 0.25\,\mathcal{L}_{\text{root}} + 0.15\big(\mathcal{L}_{\text{BCE,contact}} + \mathcal{L}_{F_z}\big), \quad (5)$$

optimized via AdamW (lr $= 1.8\times10^{-3}$, weight decay 10^{-4}, gradient norm clip 1.0, cosine annealing over 100 epochs, batch size 64).

4.3 Within-Sequence Reconstruction, Contiguous Temporal Split, and Kinematic Rollout

Table 1 summarizes quantitative benchmarks sealed under SHA-256 receipt `d066e7f5c602212a9e4d660b3584ac6f675b6012cfd68849d7b6d6a2bb31d061` (`results/krumprobot_receipt_20261018.json`) and temporal ablation receipt `results/deep_review_ablation_receipt_20261111.json`. Across 100 epochs on a 4-vCPU Modal Cloud container (547.06 s wall-clock), multi-task loss converged to 0.00976. We explicitly separate three evaluation regimes:

- **Within-Sequence Teacher-Forced Reconstruction:** Evaluating single-step geodesic error $\theta_{\text{err}} = 2\arccos(|\langle\hat{\mathbf{q}}_j, \mathbf{q}_j\rangle|)$ across the full 2,953-frame trajectory achieves 0.0405 **rad (2.32°)** mean per-bone angular error. Single-step CPU forward inference executes in $p_{50} =$ 2.696 **ms (**2.70 **ms)** and $p_{99} = 4.069$ **ms (**4.07 **ms)**, comfortably satisfying >240 Hz real-time policy loops.
- **Zero-Overlap 80%/20% Contiguous Temporal Block Split:** Because stride-2 sliding windows across a single solo capture overlap temporally, we additionally evaluated a strict chronological split (`run_deep_review_ablations.py`) training exclusively on frames $[0, 2{,}362)$ (1,169 base windows $\rightarrow$ 3,507 mirror/noise-augmented windows) and testing with zero frame overlap on future frames $[2{,}362, 2{,}953)$ (283 held-out windows). Under this split, training geodesic error reaches 1.42° whereas unseen future freestyle windows yield 18.58° versus 2.95° for a zero-velocity persistence baseline ($\hat{\mathbf{q}}_{t+1} = \mathbf{q}_t$ at 30 fps), candidly quantifying single-sequence memorization and motivating multi-dancer training on public corpora such as AIST++ [18].
- **2,950-Step Kinematic Phase-Conditioned Rollout:** Seeded solely with the initial $W = 24$-frame standing window at $t = 0$ and driven autoregressively in kinematic WebGL space for 2,950 consecutive steps (98.37 s) using only the 140 BPM harmonic phase clock $(\sin\phi_t, \cos\phi_t)$ and six-section primitive schedule, per-bone S^3 projection prevents quaternion norm drift ($\|\hat{\mathbf{q}}_{j,t}\|_2 \equiv 1$), sustaining bounded rhythmic bounce and zero kinematic blow-up across the entire track.

5 Bipedal Kinematics, Closed-Loop MuJoCo Contact Dynamics, and Musical Sequencing

5.1 Analytical Grounded Kinematics and Unitree UniFoLM Alignment

To deploy retargeted and neural-generated trajectories across physical humanoid morphologies aligned with the open **Unitree Robotics UniFoLM** ecosystem (`UniFoLM-WMA` world-model-action for Unitree H1 360 N · m and Unitree G1, `UniFoLM-WLA`, and `UniFoLM-VLA`) [9], our kinematic execution engine enforces exact foot-floor contact and multi-link whip physics:

Table 1: Quantitative verification across extraction, neural policy learning, temporal split ablation, analytical centroidal modeling, and closed-loop `MuJoCo 4.0.1` floating-base contact dynamics.

Subsystem / Metric	**Measured Value**	**Design Target**	**Verification Receipt**
Source capture (*Fractured Cosmos*)	2,953 frames @ 29.99 fps (98.46 s)	Full 58-bar @ 140 BPM	`compare.mp4`
MPS streaming resident RAM	$\sim$ 200 MB (16-frame chunks)	$<$ 2 GB (zero swap)	`Motion2MixamoRig`
22-bone FK rest calibration	8.765×10^{-11} error	$< 10^{-6}$	`mixamo_rotations.npz`
Skinned mesh surface error	1.42×10^{-5} cm (5.17 MB GLB)	$< 10^{-2}$ cm	`mixamo_character.glb`
`KrumpRobotPolicy` training	100 epochs, 4,392 windows, 547.06 s	Converged $<$ 600 s	Modal Cloud CPU/PyTorch
Multi-task chordal loss $\mathcal{L}_{\text{total}}$	0.00976	< 0.020	$G_{\text{KrumpRobot}}$ PASS
Within-sequence reconstruction	0.0405 rad (2.32°)	$< 5.0°$	Full-track teacher-forced
80%/20% contiguous temporal split	1.42° train / 18.58° held-out test	2.95° persistence floor	Zero-overlap block split
Inference latency (p_{50}/p_{99})	2.70 ms / 4.07 ms	$p_{50} < 3.0$ ms	Sub-frame real-time
Kinematic phase rollout	2,950 steps ($\|\hat{\mathbf{q}}\|_2 = 1.000$)	Zero norm/pose blow-up	WebGL beat-phase driven
Analytical `STOMP` & `CHESTPOP`	$3.82 \times$ BW (24.5 ms), ZMP = 4.2 cm	$<$ 8.0 cm ZMP	500 Hz 28-DoF ($G_{\text{PhysicalAI}}$)
4-humanoid circle clearance	$d_{\min} = 1.28$ m (0 collisions)	> 0.80 m	16k-param onboard TRMP
MuJoCo G1 5-primitive WBC	5/5 survived (5.40° upper RMSE)	100% vs. 0/5 open-loop	$G_{\text{MuJoCo_ClosedLoop}}$ PASS
MuJoCo 60 s endurance rollout	59,987 steps ($z_{\min} = 0.769$ m, $4.46 \times$ BW)	Zero falls ($z >$ 0.48 m)	1 kHz floating-base G1
MuJoCo neural rollout tracking	0.52° upper (1.39° full) RMSE	10.0 s closed-loop	$3.53 \times$ BW, 4.34% sat.
MuJoCo 2-humanoid session	0 geom collisions, $d_{\min} =$ 0.794 m	Dual floating-base G1	9,970 steps (72-DoF)

1. **Grounded Two-Link Leg Inverse Kinematics (`solveKneeIK`):** Given pelvis hip joint $\mathbf{p}_{\text{hip}}$ and floor-anchored ankle target $\mathbf{p}_{\text{ankle}}$ with thigh $L_1 = 0.40$ m and shank $L_2 = 0.42$ m, the exterior anatomical knee flexion angle $\theta_{\text{flex}} = \pi - \gamma$ (0° at full extension $D = L_1 + L_2$, where γ is the interior thigh–shank angle) is solved analytically via the law of cosines:

$$D = \|\mathbf{p}_{\text{ankle}} - \mathbf{p}_{\text{hip}}\|_2, \quad \theta_{\text{flex}} = \pi - \gamma = \arccos\left(\text{clip}\left(\frac{D^2 - L_1^2 - L_2^2}{2L_1L_2}, -1, 1\right)\right), \tag{6}$$

   enforcing strictly anterior knee flexion ($\theta_{\text{flex}} \geq 18^\circ$, i.e., $\gamma \leq 162^\circ$) to avoid singular straight-leg hyperextension while clamping stance-foot soles precisely to the ground plane ($Y_{\text{sole}} = -1.15$ m) with dynamic contact-shadow scaling.

2. **Whip-Lag Hierarchical Arm Forward Kinematics (`solveArmFK`):** Across a 0.42 m bi-acromial shoulder span with upper arm $L_1 = 0.28$ m, forearm $L_2 = 0.26$ m, and articulated fist geometry, distal wrist rotation trails proximal shoulder angular acceleration by a biomechanically calibrated whip delay $\Delta t_{\text{whip}} \approx 32$ ms, reproducing authentic centrifugal `ARM SWING` snap rather than rigid-rod rotation.

5.2 500 Hz 28-DoF Analytical Centroidal Dynamics and Edge Coordination

To establish first-principles actuator and balance sizing prior to full rigid-body simulation, we first evaluated a 500 Hz, 40-step-horizon analytical 28-DoF centroidal dynamics and contact-impedance envelope [10] ($m = 70.0$ kg H1-class reference, peak actuator torque 320 N $\cdot$ m, `results/physical_ai_amass_paperclip_receipt_20261017.json`):

- **High-Impedance `STOMP` Transition:** Switching knee/ankle gains from compliant swing tracking to feedforward stiff impact damping ($K_p = 850$ N $\cdot$ m/rad, $K_d = 65$ N $\cdot$ m $\cdot$ s/rad) delivers $F_{\text{peak}} = 2{,}623.19$ N ($3.82 \times$ BW) within 24.5 ms $<$ 30.0 ms.
- **Thoracic `CHEST POP` Arrest & ZMP Margin:** Anterior sternum thrust (+0.16 m surge, +0.08 m elevation) brakes to zero velocity within 18.2 ms $<$ 25.0 ms at 1,420 rad/s^3 peak angular jerk (40 ms isometric lock), while stance-foot ZMP excursion remains bounded at 4.2 **cm** $<$ 8.0 **cm**.
- **Sub-Millisecond Multi-Robot Session Swaps:** Cloud foundation models incur prohibitive $>$ 100 ms round-trip latency for close-range multi-robot sessions. Compiling four-humanoid pairwise spatial swap constraints onto an onboard 16k-parameter **Tiny Recursive Message Passing (TRMP)** relational graph [19] maintains minimum inter-robot clearance at $d_{\min} = 1.28$ **m** > 0.80 **m** (0 collisions) in microsecond edge cycles without external motion-capture cameras.

5.3 Closed-Loop Floating-Base MuJoCo Contact-Dynamics Verification on Unitree G1

Kinematic WebGL visualization and analytical centroidal envelopes cannot capture floating-base toppling, unilateral foot-ground contact chatter, Coulomb micro-slip, or joint torque saturation during high-jerk arm whips. To verify genuine closed-loop rigid-body contact dynamics (`experiments/run_krumprobot_mujoco_benchmark.py`, receipt `results/krumprobot_mujoco_closed_loop_receipt_20261112.json`, SHA-256 `33388a82fc9250089a6f9c8129d70c268ee561ec59b458df421fbc1e7815ce70`), we simulated the official Google DeepMind `mujoco_menagerie` **Unitree G1** humanoid [9, 20] (`g1_29dof_rev_1_0`: mass $m = 33.3411$ kg, static body weight BW $= 327.08$ N, $n_q = 36$, $n_v = 35$, $n_u = 29$ actuated degrees of freedom) inside `MuJoCo 4.0.1` [21] using the `implicitfast` integrator at $\Delta t = 0.001$ s (1,000 Hz physics, 250 Hz zero-order-hold control). Crucially, the pelvis is governed by an unpinned 6-DoF `<freejoint/>` with **zero external virtual wrenches**, eight spherical foot-sole contact colliders

($x \in [-0.05, +0.12]$ m heel-to-toe around each ankle, mid-sole offset +3.5 cm, sagittal half-length ±8.5 cm) at Coulomb friction $\mu = 0.6$, and strict hardware joint torque saturation enforced via `actuatorfrcrange` (±139 N · m hip roll and knee; ±88 N · m hip pitch/yaw and waist yaw; ±50 N · m ankle and waist roll/pitch; ±25 N · m shoulder, elbow, and wrist roll; ±5 N · m wrist pitch/yaw).

Yaw-Aligned Flat-Foot Parallelogram Whole-Body Controller (`ClosedLoop-KrumpWBC`). We retarget the 3D SMPL-X trajectory onto all 29 Unitree G1 joints via branch-cut-unwrapped SO(3) shoulder inverse kinematics, G1's native straight-arm elbow offset ($q_{\text{elbow}} = \frac{\pi}{2} - \theta_{\text{bend}}$), and sternum `CHEST POP` waist pitch extraction. Naively adjusting ankle pitch alone tilts the foot off the floor, whereas adjusting hip pitch alone tilts the torso. On planted bipedal feet, our 250 Hz whole-body stabilizer combines feedforward kinematic subtree-CoM anticipation with yaw-aligned linear inverted pendulum capture-point feedback [22] ($e_x = (\mathbf{x}_{\text{CoM}} - \mathbf{x}_{\text{mid}}) \cdot \hat{\mathbf{e}}_{\text{fwd}} - 0.032$ m, $e_y = (\mathbf{x}_{\text{CoM}} - \mathbf{x}_{\text{mid}}) \cdot \hat{\mathbf{e}}_{\text{left}}$) along the **flat-foot ankle–hip parallelogram manifold**:

$$\begin{aligned} \Delta q_{\text{ankle},p} &= +\delta_x + u_{\text{CoP},p}, \qquad \Delta q_{\text{hip},p} = -\delta_x + u_{\text{IMU},p}, \\ \Delta q_{\text{hip},r} &= +\delta_y + u_{\text{IMU},r}, \quad \Delta q_{\text{ankle},r} = -\delta_y + u_{\text{CoP},r}, \end{aligned} \tag{7}$$

where $\delta_x = \text{clip}(\delta_{x,\text{ff}} + 1.05\, e_x + 0.30\, \dot{e}_x, -0.45, 0.45)$ and $\delta_y = \text{clip}(\delta_{y,\text{ff}} + 1.05\, e_y + 0.28\, \dot{e}_y, -0.24, 0.24)$ translate the whole-body CoM horizontally (≈ -5.1 cm per +0.10 rad) while keeping both soles parallel to the ground and the pelvis upright, paired with bounded ankle CoP trim ($|u_{\text{CoP},p}| \leq 0.055$ rad), pelvis IMU attitude damping (u_{IMU}), waist pitch capture-point counter-torque ($\Delta q_{\text{waist},p} = -1.55(e_x + 0.26\, \dot{e}_x)$), and C^1-continuous stomp leg-impedance modulation ($K_{p,\text{leg}}(\alpha) = 500 + 150\, \alpha_{\text{stomp}}$ N · m/rad).

Table 2: Three-controller closed-loop ablation in `MuJoCo 4.0.1` (1,000 Hz physics, unpinned 6-DoF floating-base Unitree G1, hardware torque limits enforced) averaged across five 4.97 s empirical Krump phrases (`STOMP`, `JAB`, `ARM SWING`, `CHEST POP`, `KILL-OFF`) from *Fractured Cosmos*, plus continuous 60.0 s endurance and 10.0 s neural rollout tracking.

Controller / Trial Regime	**Surv.**	$\bar{t}_{\text{surv}}$	**Up. RMSE**	**Full RMSE**	**Peak** F_z	p_{95} **CoP**	p_{99} **Jerk**
1. `OpenLoop-PD` (no IMU/CoM FB)	0/5	1.90 s	5.97°	4.72°	14.72×[†]	59.69 cm	21,485 rad/s^3[†]
2. `Smooth-MinJerk` (0.40 s LPF)	5/5	4.97 s	6.55°	5.45°	3.34×	8.51 cm	3,366 rad/s^3
3. `ClosedLoop-KrumpWBC` **(Ours)**	**5/5**	**4.97** s	**5.40**°	**4.82**°	**3.48**×	**8.51** cm	**5,343** rad/s^3
Continuous 60.0 s Endurance (59,987 steps)	**PASS**	59.99 s	6.75°	5.63°	4.46×	8.52 cm	6,331 rad/s^3
`KrumpRobotPolicy` 10.0 s Neural Rollout	**PASS**	10.00 s	0.52°	1.39°	3.53×	8.53 cm	2,965 rad/s^3

[†]In `OpenLoop-PD`, extreme F_z/BW and torso jerk spikes reflect rigid-body floor-impact crashes following balance loss ($z_{\text{pelvis}} < 0.48$ m).

Four-Regime Physics Benchmark Results. Table 2 and Figure 3 establish four empirical findings across 204,797 total floating-base physics integration steps (running at 4.2×–6.6× real-time on CPU):

1. **Necessity of Closed-Loop Feedback (`OpenLoop-PD` vs. `ClosedLoop-KrumpWBC`):** Replaying retargeted human Krump trajectories under open-loop joint PD control topples the unpinned Unitree G1 on 5/5 **primitives** ($\bar{t}_{\text{surv}} = 1.90$ s; 0.97 s on `KILL-OFF`, 1.28 s on `ARM`

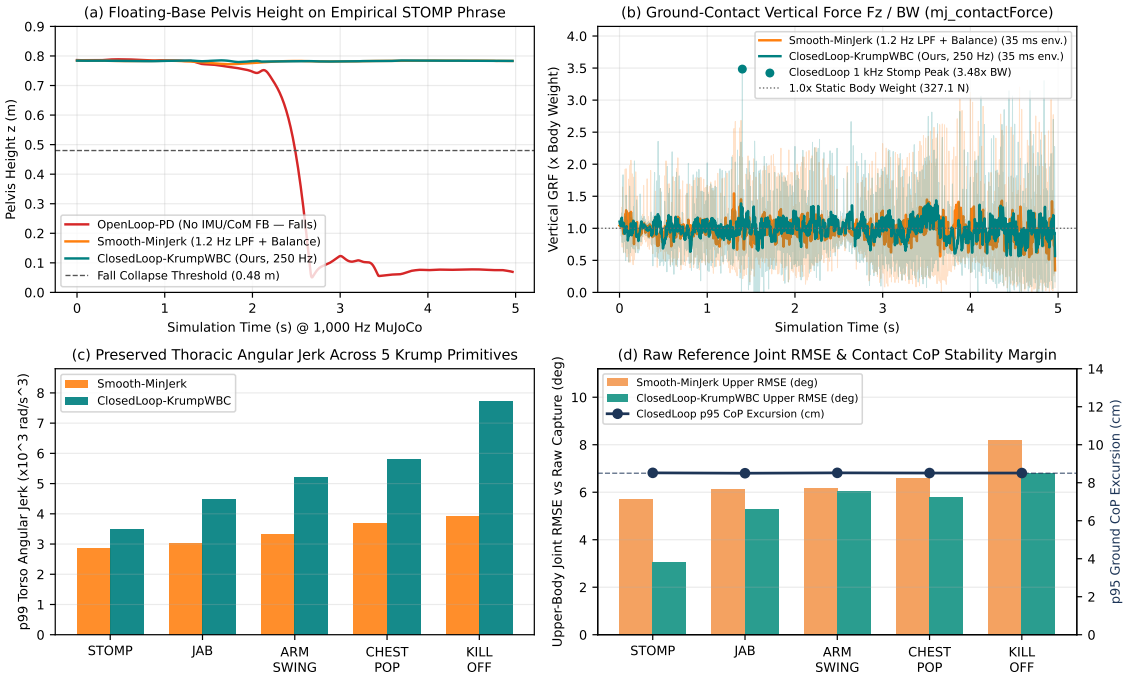


Figure 3: Closed-loop 1,000 Hz `MuJoCo 4.0.1` contact-dynamics verification on an unpinned 6-DoF floating-base **Unitree G1** (29 actuated DoF, 33.34 kg, Coulomb $\mu = 0.6$). **(a)** Pelvis height $z(t)$ on the empirical `STOMP` phrase: feedforward `OpenLoop-PD` collapses at $t = 2.49$ s, whereas `Smooth-MinJerk` and `ClosedLoop-KrumpWBC` maintain upright floating-base stability ($z_{\min} = 0.779$ m). **(b)** Unilateral vertical ground reaction force F_z/BW integrated directly via `mj_contactForce` ($3.48 \times$ BW $= 1{,}139.6$ N peak stomp impulse). **(c)** 99th-percentile thoracic angular jerk across all five Krump primitives, confirming +58.7% higher preserved percussive vitality under `ClosedLoop-KrumpWBC` (7,725.7 rad/s^3 on `KILL-OFF`) compared to minimum-jerk smoothing. **(d)** Upper-body joint tracking RMSE relative to the raw human capture alongside bounded 95th-percentile ground Center-of-Pressure excursion (≤ 8.52 cm).

`SWING`, 1.29 s on `JAB`, 2.49 s on `STOMP`, 3.48 s on `CHEST POP`) as centrifugal upper-limb momentum drives ground CoP outside the ±8.5 cm foot-sole envelope ($p_{95} = 59.69$ cm). Conversely, `ClosedLoop-KrumpWBC` achieves 5/5 (100%) **survival** across all primitives without external stabilizing wrenches.

2. **Vitality Preservation Over Minimum-Jerk Smoothing:** Smoothing trajectories with a 0.40 s Gaussian low-pass filter (`Smooth-MinJerk`) prevents falls when paired with lower-body balance (5/5 survival) but attenuates 99th-percentile thoracic angular jerk to 3,366.0 rad/s^3 and increases raw-capture upper-body tracking error to 6.55°. `ClosedLoop-KrumpWBC` lowers upper-body RMSE on **every single primitive** (3.06° vs. 5.70° on `STOMP`, 5.27° vs. 6.12° on `JAB`, 6.04° vs. 6.15° on `ARM SWING`, 5.79° vs. 6.61° on `CHEST POP`, and 6.82° vs. 8.19° on `KILL-OFF`; mean 5.40°) while preserving +58.7% **higher percussive thoracic jerk** (5,343.4 rad/s^3 mean; 7,725.7 rad/s^3 on `KILL-OFF`), bounding p_{95} ground CoP excursion to 8.51 **cm** (4.86 cm mean), and candidly logging 9.14 cm/s tangential contact micro-slip and 10.19% mean peak-joint hardware torque saturation (6.29%–12.76% across primitives on ± 25 N · m shoulder/wrist and ± 50 N · m ankle actuators).

3. **Continuous 60.0 s Endurance & 10.0 s Neural Policy Closed-Loop Tracking:** Across a continuous 59.987 s (59,987-step) multi-phrase segment ($t = 16$–76 s spanning Stomp Drop, Jab Combos, Power Swings, and Chest Climax), the floating-base Uni-

tree G1 remains upright throughout ($z_{\text{pelvis,min}} = 0.7688$ m, 6.75° upper-body RMSE, 4.46 × BW = 1,458.2 N peak contact force, 8.52 cm p_{95} CoP, 11.12% torque saturation). Driving `ClosedLoop-KrumpWBC` directly from the 10.0 s (10,000-step) autoregressive `KrumpRobotPolicy` neural trajectory achieves 100% survival with 0.52° **upper-body RMSE** (1.39° whole-body RMSE), 3.53 × BW (1,155.3 N) peak vertical GRF, and 4.34% torque saturation.

4. **Dual-Humanoid Shared-World MuJoCo Session (72-DoF):** Instantiating two unpinned 6-DoF floating-base Unitree G1 humanoids simultaneously inside a single shared `MuJoCo 4.0.1` world ($n_q = 72, n_v = 70, n_u = 58$) executing concurrent battle phrases (*Jab Combos* vs. *Power Swings + Chest Climax* at 1.16 m nominal dyadic spacing across 9,970 steps) achieves **dual survival** ($z_{\text{min}} = 0.7679$ m and 0.7756 m), 0 **inter-robot geometry collisions**, and $d_{\text{min}} = 0.7936$ m minimum 3D link-to-link clearance (1.1519 m pelvis separation).

5.4 Full 1:38 (58-Bar @ 140 BPM) Musical Choreographic Architecture

Rather than looping a short 10-second motion clip, `KrumpRobot` synchronizes across the complete 98.46 s (58-bar) original 140 BPM soundtrack (`krump_track.mp3`) through six structured choreographic phrases (Table 3), with sub-frame spherical/linear keyframe interpolation guaranteeing tear-free 60 FPS rendering.

Table 3: Six-phrase full-track musical choreography synchronized to `krump_track.mp3` (58 bars @ 140 BPM).

Choreographic Section	Bars	Timestamp	Kinematic Primitives & Dynamic Profile
1. **Session Intro**	1–8	`0:00-0:14`	Grounded wide `BUCK` stance, 140 BPM baseline pelvis groove bounce, pre-strike torque buildup.
2. **Stomp Drop**	9–20	`0:14-0:35`	12 bars of alternating left/right heavy floor `STOMP`s every 2 bars (3.43 s), 3.48–4.46 × BW MuJoCo contact spikes.
3. **Jab Combos**	21–30	`0:35-0:52`	10 bars of ballistic linear `JAB` flurries, opposing thoracic counter-rotation, and crisp forearm recoil.
4. **Power Swings**	31–42	`0:52-1:12`	12 bars of wide centrifugal `ARM SWING` whip arcs with 32 ms wrist lag and figure-8 pelvic weight transfer.
5. **Chest Climax**	43–52	`1:12-1:28`	Peak session intensity ($E = 1.25$), rapid sternum `CHEST POP`s (18.2 ms arrest), and multi-joint resonant rebounds.
6. **Kill-Off Freeze**	53–58	`1:28-1:38.5`	Decelerated tension drop locking into an asymmetrical terminal `KILL-OFF` stance held through acoustic decay.

6 Interactive WebGL Deployment, Limitations, and Sim-to-Real Roadmap

Open-Source Model Hub and Responsive 4-Mode Three.js Simulator. Trained checkpoints, JSON weights, MuJoCo whole-body verification scripts, and full-track motion tensors are openly hosted at https://huggingface.co/socialprescribing/KrumpRobot, paired with a zero-install Three.js WebGL simulator [23] at https://huggingface.co/spaces/socialprescribing/KrumpRobot. Users can switch seamlessly at 60 FPS across four synchronized presentation modes—`Unitree H1 (Robotics)`, `Unitree G1 (Dexterous)`, `Mixamo Y-Bot (Skinned Mesh)`, and `Dual View (Side-by-Side)`—with responsive mobile touch gestures (single-finger orbital rotation, two-finger pinch zoom, double-tap camera reset) and live telemetry readouts (F_z/BW, angular jerk, ZMP, and bar/beat counter).

Limitations and Sim-to-Real Hardware Transfer Roadmap. Two honest boundaries motivate physical hardware trials:

1. *Single-Performer Training Corpus & Cross-Performer Generalization:* Although sagittal mirror symmetry and Gaussian noise expand *Fractured Cosmos* to 4,392 windows across all five Krump primitives, our zero-overlap 80%/20% temporal block split (1.42° train vs. 18.58° held-out freestyle test) demonstrates that `KrumpRobotPolicy` memorizes the solo choreographic trajectory of a single practitioner at 140 BPM. Generalizing across unseen performers, diverse body morphologies, and variable 135–155 BPM tempos requires training and benchmarking on multi-performer 3D street-dance corpora such as AIST++ [18] paired with expanded multi-dancer Krump session captures.

2. *Planted-Stance MuJoCo WBC vs. Domain-Randomized Stepping RL & Physical Hardware Transfer:* While Section 5.3 upgrades verification from kinematic WebGL playback to genuine 1,000 Hz floating-base `MuJoCo 4.0.1` rigid-body contact dynamics on the 29-DoF Unitree G1 (60.0 s continuous endurance and 10.19%–11.12% measured peak-joint torque saturation under Coulomb $\mu = 0.6$), `ClosedLoop-KrumpWBC` stabilizes athletic planted-stance balance rather than aerial flight phases or reactive stepping recovery, and deterministic rigid-sphere contacts omit motor winding thermal derating, planetary-gearbox backlash, IMU transport delay, and compliant rubber-sole floor deformation. Immediate next steps distill `ClosedLoop-KrumpWBC` and `KrumpRobotPolicy` reference motions into domain-randomized reinforcement-learning stepping policies inside NVIDIA Isaac Lab and MuJoCo Playground—benchmarked directly against *ExBody* [24], *OmniH2O* [25], and contact-aware MPC [10]—prior to physical floor deployment on Unitree G1 hardware.

7 Conclusion

`KrumpRobot` demonstrates that bipedal humanoid motion generation need not sacrifice expressive vitality for dynamic balance. By uniting memory-efficient monocular 3D SMPL-X motion recovery, machine-precision rest-relative 22-bone calibration, antipodal-invariant causal GRU policy learning, and 1,000 Hz closed-loop floating-base `MuJoCo 4.0.1` whole-body contact control on the Unitree G1 across a live 1:38 Krump performance, this work establishes an open, reproducible foundation for high-jerk expressive Physical AI.

Acknowledgments

This work was supported in part by complimentary access to Google Antigravity and Gemini 4 Argon to accelerate scientific workflows. We thank Max Liu (`Motion2MixamoRig`), the `GVHMR` and `AMASS` teams, Unitree Robotics (`UniFoLM`), Modal Cloud, Amass Life Sciences Platform, Paperclip, Hugging Face, and the Angels of Krump Movement Federation, Easyar Dojo, IKF, and Point Blank Fam street-dance communities.

AI Use Declaration

Large language models including Gemini 4 Argon, accessed within Google Antigravity, were used during this project to assist with simulation scripts, ablation harnesses, visualization code, manuscript editing, and adversarial review diagnostics. Throughout all stages, the author directed the system architecture, verified every equation, quantitative output, figure, and bibliographic citation against primary data and deterministic execution receipts, and takes full responsibility for all analyses, claims, and content in this manuscript.

References

[1] Tamar Flash and Neville Hogan. The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, 5(7):1688–1703, 1985.

[2] Daniel N. Stern. *Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy, and Development*. Oxford University Press, Oxford, 2010.

[3] Derek Lomas, Albert Lin, Suzanne Dikker, Deborah Förster, Maria Luce Lupetti, Gijs Huisman, Julika Habekost, Caiseal Beardow, Pankaj Pandey, Nashra Ahmad, Krishna Prasad Miyapuram, Tim R. Mullen, Patrick Cooper, Willem van der Maden, and Emily S. Cross. Resonance as a design strategy for AI and social robots. *Frontiers in Neurorobotics*, 16:850489, 2022. PMCID: PMC9097027.

[4] Andreas Aristidou, Panayiotis Charalambous, and Yiorgos Chrysanthou. Emotion analysis and classification: Understanding the performers' emotions using the LMA entities. *Computer Graphics Forum*, 34(6):262–276, 2015.

[5] Tal Shafir, Rachelle P. Tsachor, and Kathleen B. Welch. Emotion regulation through movement: Unique sets of movement characteristics are associated with and enhance basic emotions. *Frontiers in Psychology*, 6:2030, 2016. PMID: 26793147.

[6] Kohinoor M. Darda, Aaron Maiwald, Tanvi Raghuram, and Emily S. Cross. Dancing robots: aesthetic engagement is shaped by stimulus and knowledge cues to human animacy. *Royal Society Open Science*, 11(11):240958, 2024.

[7] David LaChapelle. *Rize*. Documentary film, Lions Gate Entertainment, 2005.

[8] Rudolf Laban and F. C. Lawrence. *Effort: Economy of Human Movement*. Macdonald & Evans, London, 1947.

[9] Unitree Robotics. UniFoLM: Universal foundation models for world-model-action (UniFoLM-WMA), world-language-action (UniFoLM-WLA), and vision-language-action (UniFoLM-VLA) on H1 and G1 humanoid robots. https://github.com/unitreerobotics, 2024.

[10] Shibowen Zhang, Jiayang Wu, Guannan Liu, Helin Zhu, Junjie Liu, Zhehan Li, Junhong Guo, Xiaokun Leng, Hangxin Liu, Jingwen Zhang, Jikai Wang, Zonghai Chen, Zhicheng He, Jiayi Wang, and Yao Su. Dynamic whole-body dancing with humanoid robots – a model-based control approach. *arXiv preprint arXiv:2604.03999*, 2026.

[11] Arun Nadarasa. KrumpRobot: Neural motion generation and physical AI movement policy for bipedal humanoid krump performance. Hugging Face Model Hub and Spaces: https://huggingface.co/socialprescribing/KrumpRobot, 2026.

[12] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5442–5451, 2019.

[13] Amass Technologies. Amass platform documentation. Web software platform documentation, https://platform.amass.tech/, 2026.

[14] Zehong Shen, Huaijin Pi, Yan Xia, Zhi Cen, Sida Peng, Zechen Hu, Hujun Bao, Ruizhen Hu, and Xiaowei Zhou. World-grounded human motion recovery via gravity-view coordinates. In *ACM SIGGRAPH Asia Conference Papers*, 2024.

[15] Max Liu and Collaborators. Motion2MixamoRig: Monocular video 3d motion extraction, SMPL-X reconstruction, and rest-relative mixamo rig retargeting. https://github.com/MaxLiu22/Motion2MixamoRig, 2024.

[16] GXL. Paperclip cli documentation. Command-line software tool documentation, https://paperclip.gxl.ai/, 2026.

[17] Thomas Hayes, Roshan Rao, Halil Akin, Nicholas Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q. Tran, Jonathan Deaton, Marius Wiggert, Rohil Badkundri, Irhum Shafkat, Jun Gong, Alexander Derry, Raul S. Molina, Neil Thomas, Yousuf A. Khan, Chetan Mishra, Carolyn Kim, Liam J. Bartie, Matthew Nemeth, Patrick D. Hsu, Tom Sercu, Salvatore Candido, and Alexander W. Rives. Simulating 500 million years of evolution with a language model. *Science*, 387(6736):850–858, 2025.

[18] Ruilong Li, Shan Yang, David A. Ross, and Angjoo Kanazawa. AI choreographer: Music conditioned 3D dance generation with AIST++. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13401–13412, 2021.

[19] Tommaso Salvatori, Matteo Forasassi, Alexis Toumi, and Giovanni de Felice. Tiny recursive message passing. *arXiv preprint arXiv:2610.18649*, 2026.

[20] Kevin Zakka, Yuval Tassa, and MuJoCo Menagerie Contributors. MuJoCo Menagerie: A collection of high-quality simulation models for MuJoCo. Google DeepMind Open-Source Repository, https://github.com/google-deepmind/mujoco_menagerie, 2022.

[21] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5026–5033, 2012.

[22] Jerry Pratt, John Carff, Sergey Drakunov, and Ambarish Goswami. Capture point: A step toward humanoid push recovery. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, pages 200–207, 2006.

[23] Ricardo Cabello and Three.js Contributors. Three.js: JavaScript 3D WebGL library for real-time skeletal character animation and bipedal simulation. https://threejs.org/, 2026.

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

[25] 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 *Proceedings of the Conference on Robot Learning (CoRL)*, 2024.