Summarization: Chasing Stability: Humanoid Running via Control Lyapunov Function Guided Reinforcement Learning

Summarized by Xiaonan (Nice) Wang *
Summarization† Generated based on Paper from Olkin et al. (2026)
Topic: Control Lyapunov Functions, Humanoid Running, Reinforcement Learning, Sim-to-Real

February 11, 2026

Abstract

This document summarizes the core contributions and methodology of the paper "Chasing Stability: Humanoid Running via Control Lyapunov Function Guided Reinforcement Learning, Olkin et al. [2025]", focusing on its' main ideas and the core blocks.

1 Overview: Core Questions and Answers

(1) What is the problem?

– **Nonlinear** and **hybrid** system for **humanoid running**.

(2) Why need to solve this problem?

– Achieving performant and robust running;

– Rejecting disturbances.

(3) How is it different from prev.?

– Embed multi-domain trajectories and CLFs directly into the RL reward, **no** trajectories or CLFs at **runtime**;

– Achieve **certifiable stability**.

(4) Why is it better than prev.? (Advantages)

– Compared with *heuristically designed controller*: can succeed in **Hybrid Zero Dynamics (HZD)**;

– Compared with *offline trajectory optimization + online tracking with traditional feedback linearization & control Lyapunov functions (CLFs)*: operate on both **continuous** and **discrete** (i.e. hybrid) dynamics, with capacity to generate **transient** behaviors;

– Compared with *(contact) impact invariance* solutions like *MPC*: yield humanoid running and **computational efficient**;

– Compared with *end-to-end RL*: **eliminate** the need to **handcraft** and **tune heuristic** reward terms;

– Produce **steady state** running motions with accurate position and velocity tracking.

*wangxiaonannice@gmail.com
†**Disclaimer:** This summarization is for research and study purposes. It represents a personal interpretation and may contain inaccuracies. Feedback or corrections via email are highly appreciated.

(5) What is the approach itself?

– **Sim-to-Real Hybrid Controller (System)** $\mathcal{H} = (\mathcal{D}, \mathcal{S}, \Gamma, \Delta, \mathcal{F})$ for **Humanoid Running**

- Robust and high-performance **policies** for humanoid running.

– Controller Synthesizing (i.e. Policy Learning) Framework – **CLF-RL**:

- **Offline Multi-Domain Trajectory Optimization**
  - Domain: (*SSP*, *FLT*)
  - Optimization (*Eq. 10*) (subject to):
    - Hybrid Systems (*Eq. 1 (Continuous Evolution)* + *Eq. 2 (Discrete Transition)*)
    - **HZD Constraint** (*Eq. 7* or *Eq. 10e* + *Virtual Constraint Eq. 10f*)
    - State and Control Limits (*Eq. 10g*, *Eq. 10h*)
  - Output:
    - ecision variable α (Bézier curve coefficient);
    - **Reference** trajectories (i.e. **Gait Library**): $x^*(t)$ (state) and $u^*(t)$ (control);
    - **Domain timing**: time of each domain
  - Tricks and Effects:
    - HZD Constraint (*Eq. 7*):
      - **Impact Invariance**, **Periodic Stability**
    - Virtual Constraint:
      - Fitting Bézier curve enables output easy to track. -> **Trade-off** between HZD and subsequent RL;
      - **Position-based** virtual constraint -> Build *Local Reference Frame*, simplify computation, adapt to FLT;
    - **Warm Start** (fast at the beginning, then slow down):
      - Conform to **physical intuition** and guarantee **convergence**;
      - Prevent *neighboring* gaits from *converging to drastically different solutions*.
    - **Even Division** (*Same Node Amount* of Each Domain Time)
      - Ensure the fixed dimension of the optimization, facilitating calculation by solvers (eg. IPOPT).
- **Reinforcement Learning (RL)**
  - **Reward**:
    - **Tracking Reward** (*Eq. 13*):
      - Minimize *energy & error*;
      - The "deep coupling" between HZD and RL.
    - **CLF Decay & Penalty** (*Eq. 14*) (*key novelty*) for **Stability**:
      - CLF Constraint (*Eq. 8*) enables **exponentially stabilizing**;
      - **Soft Penalty**: From *"Discrete Discrimination"* (*Eq. 8*) to *"Continuous Guidance"* (*Eq. 14*);
      - Hinge-like.
    - Other Rewards & Penalties: SSP Domain Reward, FLT Contact Forces Penalty, L2 & Hinge-like Regularization Penalties.
  - Other Design Tricks:
    - Robustness to **Disturbances**: RL learning with domain **randomization**, facilitating sim-to-real transfer (avoiding model *mismatch* and preventing *over exploitation*);
    - Straight Running: **Layered** Control Architecture with extra independent PD Controllers.

– Novelty:

- Soft CLF Decay in Reward (*Core*)
- The "deep coupling" between HZD and RL
- Warm Start

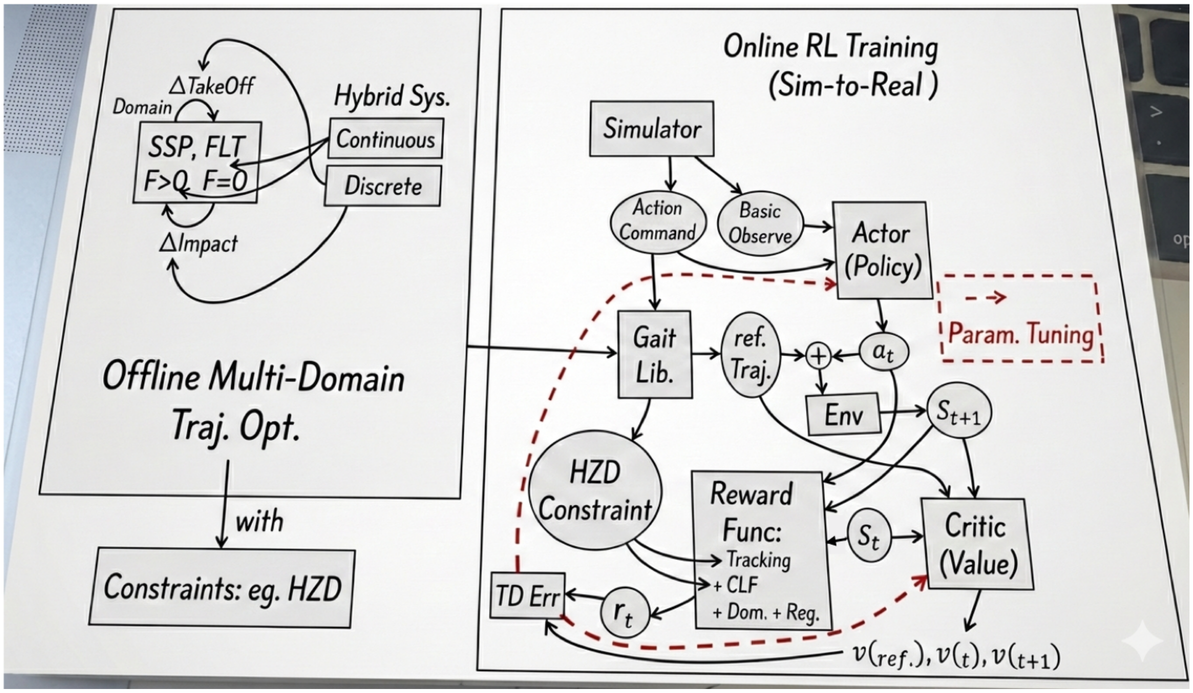

Figure 1: Summarized Block-Diagram of CLF-RL Controller Synthesizing Framework

(6) What are the applications of it?

– A **Hybrid Controller** that can be used both in simulation and real-world **Humanoid** Robots (eg. UniTree G1, Atlas) for **periodic steady state** *running*.

2 The Structure

Summarized Block-Diagram The summarized block-diagram of *CLF-RL controller synthesizing framework* see fig. 1[1].

Original in Paper The original block-diagram see fig. 2.

3 Open Questions

1. **Optimization Target of Multi-Domain Traj. Opt.:** What are the optimization targets of Multi-Domain Trajectory Optimization? Do different optimization target types affect the overall performance of the framework?

2. **Generalizing to Non-Periodic-Stable Dynamic Motions:** Can this framework generalize to non-periodic-stable dynamic motions?

References

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

[1]For efficiency, this diagram was initially hand-drawn on paper and then converted into its current digital version using Nano Banana Pro.

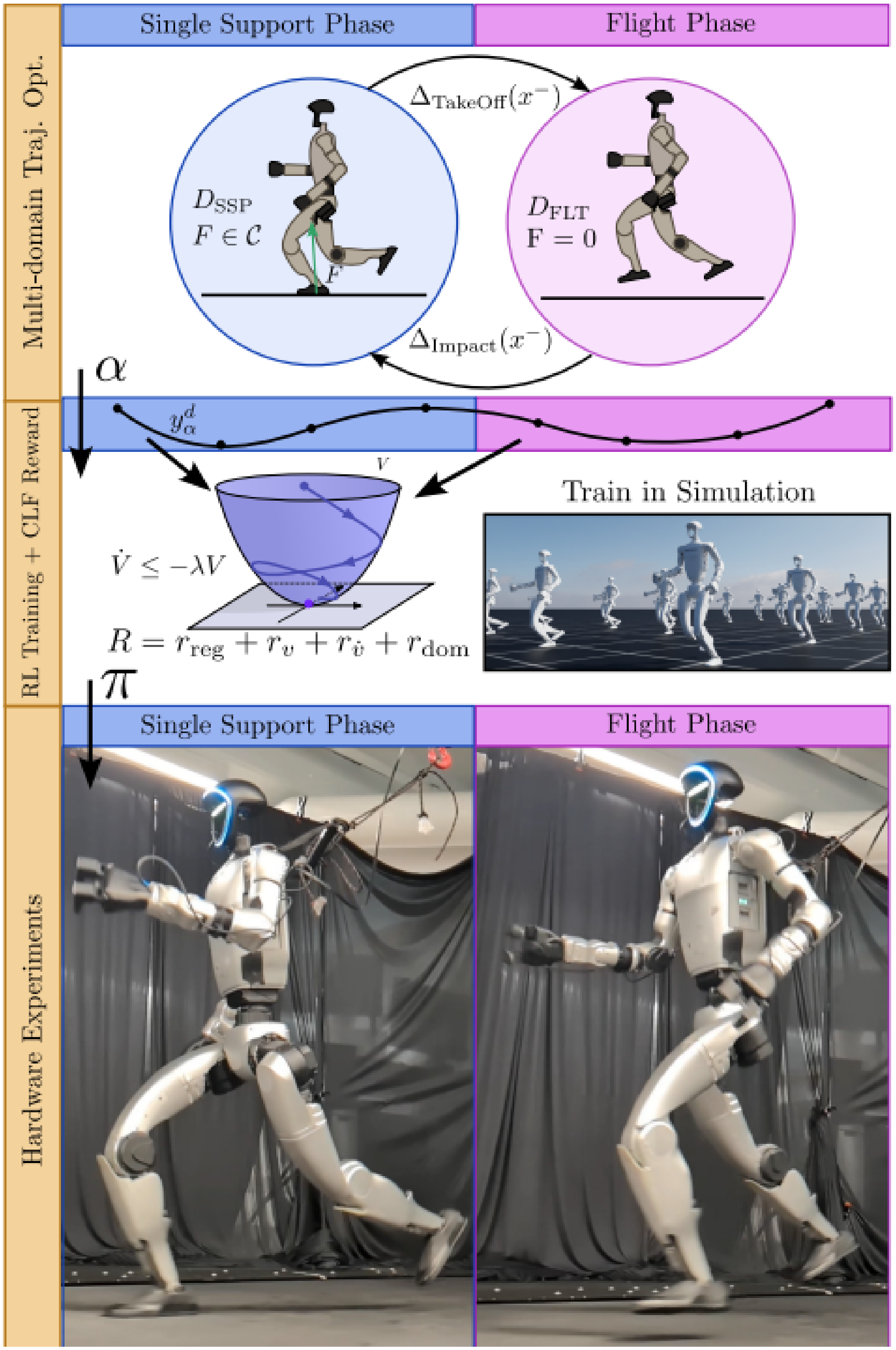

Figure 2: Original Block-Diagram in Paper