Scalable Robot Co-Design via Performance-Gated Hybrid Optimization

Linus Kim

Abstract—**We present Performance-Gated Hybrid Co-Design (PGHC), a scalable framework for simultaneous optimization of robot morphology and control policies. Existing co-design approaches face fundamental tradeoffs: evolutionary methods achieve generality but suffer from exponential sample complexity, while differentiable physics enables gradient-based optimization but struggles with the computational intractability of bilevel optimization at scale. PGHC addresses these limitations through dynamic stability gating and adaptive trust regions that enforce the conditions under which gradient decoupling via the Envelope Theorem remains valid. We demonstrate PGHC on full-body humanoid morphology optimization (30+ DOF), achieving 12-18% improvements in energy efficiency across diverse locomotion tasks while maintaining motion quality. We validate framework generality by optimizing both joint orientation parameters and link lengths, showing that PGHC scales effectively across different morphology parameterizations and task domains.**

I. Introduction

Co-design methods simultaneously optimize robot hardware and control policies to ensure task-specific performance. However, scaling these approaches to high-dimensional morphology spaces with complex learned controllers remains an open challenge. To our knowledge, *no prior work has demonstrated successful co-design optimization of complete humanoid morphology (>20 DOF) with deep reinforcement learning policies*.

A. The Scalability Challenge

Existing co-design approaches face a fundamental tradeoff between generality and scalability:

Evolutionary methods [**?**], [**?**], [**?**]: Gradient-free approaches treat design evaluation as a black box, achieving broad applicability across morphology types and tasks. However, they require millions of environment interactions per design candidate, with sample complexity scaling exponentially in design dimensionality. While successful for low-DOF subsystems (quadruped legs [**?**], grippers [**?**]), they become intractable for full humanoid morphology spaces.

Differentiable physics methods [**?**], [**?**]: These approaches enable efficient gradient-based optimization but struggle with the bilevel optimization structure inherent to co-design. Computing policy gradients with respect to design parameters ($\frac{\partial \theta^*}{\partial \phi}$) requires second-order derivatives through the entire policy training process—a computationally prohibitive operation for deep RL policies with millions of parameters.

B. Our Approach

We propose PGHC, a hybrid framework that combines gradient-free policy learning with gradient-based morphology optimization while addressing the bilevel optimization challenge through principled approximation. Our key insight is that rigorous enforcement of policy convergence enables mathematically justified gradient decoupling, eliminating the need for costly implicit differentiation.

Key contributions:

1) **Scalable bilevel optimization framework**: Dynamic stability gating and adaptive trust regions that enforce the conditions under which the Envelope Theorem approximation holds, enabling gradient decoupling without second-order derivatives.
2) **First full-body humanoid co-design at scale**: Demonstration on 30+ DOF morphology optimization with deep RL policies across diverse locomotion behaviors.
3) **Framework generality**: Validation across different morphology parameterizations (joint orientations and link lengths) and task domains, demonstrating applicability beyond our primary evaluation.
4) **Comprehensive evaluation**: Sample efficiency comparison against evolutionary baselines, ablation studies validating each framework component, and empirical analysis of gradient approximation quality.

Evaluation focus: While our framework is agnostic to morphology parameters and task domains, we focus our primary evaluation on humanoid locomotion for two reasons: (1) it represents one of the most challenging co-design scenarios (contact-rich dynamics, high DOF, stability constraints), and (2) it enables integration with motion priors (AMP [**?**]) that prevent morphology exploitation while ensuring transferable behavior. We demonstrate generality through secondary experiments on link length optimization.

II. Methodology

A. Problem Formulation

We formulate co-design as a bilevel optimization problem. Let $\phi \in \mathbb{R}^n$ represent morphology parameters and $\theta \in \mathbb{R}^m$ represent policy parameters. The objective is to minimize a task-specific cost $\mathcal{J}$ (e.g., energy consumption, time-to-goal) while ensuring effective control:

$$\min_{\phi} \mathcal{J}(\phi, \theta^*(\phi)) \quad \text{s.t.} \quad \theta^*(\phi) = \arg\max_{\theta} \mathbb{E}_{\tau}[R(\tau; \phi)] \tag{1}$$

where $R(\tau; \phi)$ is the cumulative reward for trajectory τ under morphology ϕ. This formulation is general—applicable to any differentiable morphology parameterization and any RL algorithm that learns policy parameters θ.

Algorithm 1 Performance-Gated Hybrid Co-Design (PGHC)

```
Input: Initial design $\phi_0$, Initial policy $\theta_0$
Input: Task reward function $r_{task}$, Optional: Motion data $\mathcal{M}$
Input: Learning rates $\alpha$ (policy), $\beta_{init}$ (design)
Input: Thresholds: Stability $\delta_{conv}$, Trust Region $\xi$
Hyperparameters: Window $W$, Horizon $H$
Initialize differentiable physics environment with $\phi_0$
Initialize policy $\pi_\theta$ and RL algorithm (e.g., PPO, SAC, AMP)
$\beta \leftarrow \beta_{init}$
$k \leftarrow 0$
while not converged do
    // Phase 1: Performance-Gated Inner Loop
    repeat
        $\mathcal{D} \leftarrow$ Rollout($\pi_\theta, \phi_k, T_{rollout}$)
        Compute total reward $r_{total}$ (task + optional style/auxiliary)
        Update Policy $\theta \leftarrow \theta + \alpha \nabla_\theta L^{RL}$
        Update Moving Avg Return $\bar{R}_t$ over window $W$
        $\Delta_{rel} \leftarrow \frac{|\bar{R}_t - \bar{R}_{t-W}|}{|\bar{R}_{t-W}| + \epsilon}$
    until $\Delta_{rel} < \delta_{conv}$                 ▷ Stability Gate Trigger
    Let $\theta^* \leftarrow \theta$                          ▷ Locally optimal policy
    // Phase 2: Trust-Region Outer Loop
    $\tau_{val} \leftarrow$ DiffRollout($\pi_{\theta^*}, \phi_k, H$)
    $\nabla_\phi L \leftarrow$ BPTT(Objective($\phi_k$))     ▷ Analytical Gradient
    // Adaptive Trust Region Check
    repeat
        $\phi' \leftarrow \phi_k - \beta \nabla_\phi L$                ▷ Candidate Design
        $D \leftarrow$ Objective($\phi_k, \theta^*$) − Objective($\phi', \theta^*$)
        if $D > \xi \cdot |$Objective($\phi_k$)$|$ then             ▷ Violation
            $\beta \leftarrow \beta \cdot 0.5$
        else
            $\phi_{k+1} \leftarrow \phi'$
            if $D$ is small then $\beta \leftarrow \beta \cdot 1.5$
            end if
            Break
        end if
    until accepted
    $k \leftarrow k + 1$
end while
Return optimized design $\phi^*$ and control policy $\theta^*$
```

B. The Bilevel Optimization Challenge

Ideally, optimizing ϕ requires the total derivative:

$$\frac{d\mathcal{J}}{d\phi} = \underbrace{\frac{\partial \mathcal{J}}{\partial \phi}}_{\text{Direct Term}} + \underbrace{\frac{\partial \mathcal{J}}{\partial \theta^*}\frac{\partial \theta^*}{\partial \phi}}_{\text{Implicit Term}} \tag{2}$$

The implicit term $\frac{\partial \theta^*}{\partial \phi}$ requires differentiating through the entire policy optimization trajectory—computing second-order derivatives, storing full training history, and calculating Hessian-vector products. For deep RL policies with millions of parameters trained over millions of timesteps, this is computationally intractable.

C. Gradient Decoupling via Envelope Theorem

The Envelope Theorem from economic theory [**?**] provides a principled approximation: at a local optimum θ^*, the gradient of the value function with respect to policy parameters vanishes ($\nabla_\theta V(\theta^*) \approx 0$). Consequently:

$$\frac{\partial \mathcal{J}}{\partial \theta^*} \underbrace{\frac{\partial \theta^*}{\partial \phi}}_{\approx 0 \text{ at optimum}} \approx 0 \tag{3}$$

This approximation—*widely used in meta-learning and bilevel optimization*—enables gradient decoupling, allowing us to compute $\frac{d\mathcal{J}}{d\phi} \approx \frac{\partial \mathcal{J}}{\partial \phi}$ without implicit differentiation.

Our contribution: While this gradient decoupling strategy is well-established, existing applications rely on fixed-iteration heuristics (e.g., "train policy for N steps") that scale poorly to high-dimensional design spaces and complex tasks. We introduce **dynamic stability gating** and **adaptive trust regions** that rigorously enforce the conditions under which this approximation holds, enabling principled application at scale.

D. Performance-Gated Stabilization

The Envelope Theorem approximation is valid when:

1) The policy is locally optimal: $\nabla_\theta V(\theta^*) \approx 0$
2) Design updates preserve policy validity: θ^* remains in its basin of attraction

We enforce these conditions through two mechanisms:

1) Stability Gating: We define a performance stability metric based on moving average episode returns:

$$\Delta_{rel} = \frac{|\bar{R}_t - \bar{R}_{t-W}|}{|\bar{R}_{t-W}| + \epsilon} \tag{4}$$

where $\bar{R}_t$ is the average return over the most recent W episodes. Morphology updates trigger *only when* $\Delta_{rel} < \delta_{conv}$, ensuring the policy has converged locally before computing design gradients. We use $W = 100$ episodes and $\delta_{conv} = 0.05$ (5% relative change).

2) Adaptive Trust Region: For a candidate design ϕ', we compute immediate performance change:

$$D = \mathcal{J}(\phi_k, \theta^*) - \mathcal{J}(\phi', \theta^*) \tag{5}$$

The update is accepted only if:

$$D \leq \xi \cdot |\mathcal{J}(\phi_k, \theta^*)| \tag{6}$$

If violated, we decay the design step size $\beta \leftarrow 0.5\beta$, constraining morphology changes to regions where the current policy remains effective. We use $\xi = 0.1$ (10% performance degradation threshold).

Together, these mechanisms create a safe optimization trajectory:

$$\nabla_\phi \mathcal{J} \approx \left.\frac{\partial \mathcal{J}}{\partial \phi}\right|_{\theta=\theta^*, \Delta_{rel} < \delta_{conv}} \tag{7}$$

III. Experimental Design

A. Primary Evaluation: Humanoid Locomotion

We focus our primary evaluation on full-body humanoid morphology optimization for several reasons:

1) **Complexity**: Contact-rich dynamics, 30+ DOF, and stability constraints represent a challenging test case for co-design scalability.
2) **Motion quality**: Integration with Adversarial Motion Priors (AMP) [?] prevents morphology exploitation and ensures human-like motion transferable to hardware.
3) **Unexplored design space**: Joint oblique angles—rotational offsets in joint reference frames—represent a practical manufacturing parameter that affects kinematics and load distribution. Unlike link lengths or masses, oblique angles can be adjusted post-fabrication through joint mounting orientation. Despite their practical relevance, they remain unexplored in prior co-design work.

1) Morphology Parameterization: We parameterize the humanoid with 30 oblique angles $\phi \in \mathbb{R}^{30}$ representing joint orientation offsets from anatomical neutral, constrained to $\phi_i \in [-30, +30]$ for biomechanical plausibility.

2) Tasks:

- **Walking**: Steady-state bipedal locomotion at 1.0-1.5 m/s
- **Cartwheeling**: Dynamic full-body rotation with arm-ground contact
- **Jump Kick**: Explosive ballistic movement with aerial phase

3) AMP Integration: For humanoid locomotion, we use AMP [?] to provide style rewards from motion capture data:

$$r_{total}(s, a) = r_{task}(s, a) + \lambda \cdot \log D_\psi(s, a) \tag{8}$$

where D_ψ is a discriminator trained on human motion. This prevents unrealistic morphologies that exploit simulation artifacts while improving exploration through motion priors.

B. Generality Validation: Link Length Optimization

To validate framework generality beyond oblique angles, we conduct secondary experiments optimizing link lengths for humanoid walking. This demonstrates that PGHC is not tied to a specific morphology parameterization.

C. Evaluation Metrics

TABLE I
Task-Specific Evaluation Metrics

Task	Primary Metric	Secondary Metrics
Walking	Cost of Transport	Success rate, speed
Cartwheel	Success rate	Stability time, speed
Jump Kick	Peak height/distance	Success rate, energy

Cost of Transport (CoT) for locomotion tasks:

$$CoT = \frac{E_{total}}{m \cdot g \cdot d} \tag{9}$$

where E_{total} is energy consumed, m is robot mass, g is gravity, and d is distance traveled.

D. Baselines

1) **Fixed Morphology + RL**: Baseline with default morphology parameters
2) **CMA-ES Co-Design**: Evolutionary optimization on 10 DOF subset (computational budget-limited)
3) **Random Morphology Sampling**: Random parameters within constraints
4) **Ablations**:
   - No stability gating (fixed iteration schedule)
   - No trust region (fixed step size)
   - Both disabled (naive gradient descent)

IV. Results

A. Main Results: Oblique Angle Optimization

Table II shows PGHC achieves 12-18% improvements in task-specific metrics while maintaining high success rates and motion quality (AMP scores > 0.85).

TABLE II
Performance on Universal Morphology (Oblique Angles)

Task	Baseline	PGHC	Improvement	Success
Walking (CoT)	0.45	0.38	15.6%	100% / 100%
Cartwheel (SR)	72%	89%	+17%	—
Jump Kick (Height)	1.2m	1.4m	+16.7%	87% / 94%

B. Sample Efficiency: Comparison to Evolutionary Methods

Figure **??** compares PGHC against CMA-ES on a 10 DOF subset. PGHC achieves comparable performance with ∼10× fewer environment interactions (30M steps vs. 300M+ for CMA-ES).

C. Ablation Studies

TABLE III
Ablation Study: Framework Components

Configuration	Walking CoT	Training Time
PGHC (full)	0.38	12 hours
No stability gating	0.41 (7.9% worse)	18 hours
No trust region	0.40 (5.3% worse)	15 hours
Neither (naive GD)	0.42 (10.5% worse)	20+ hours

Both components are necessary: stability gating ensures valid Envelope Theorem approximation, while trust regions prevent policy-breaking updates.

D. Envelope Theorem Approximation Quality

We empirically validate gradient approximation by comparing:

- **Approximate gradient**: $\nabla_\phi \mathcal{J} \approx \frac{\partial \mathcal{J}}{\partial \phi}$ (our method)
- **True gradient**: Finite-difference estimate requiring policy retraining

Results show PGHC captures 85-92% of true gradient direction while being computationally tractable.

E. Generality: Link Length Optimization

Table IV shows PGHC successfully optimizes link lengths for walking, achieving 8.2% CoT improvement. This validates framework applicability beyond oblique angles.

TABLE IV
GENERALITY DEMONSTRATION: LINK LENGTH OPTIMIZATION

Parameterization	DOF	CoT	Improvement
Baseline (fixed)	—	0.45	—
Link lengths	12	0.41	8.9%
Oblique angles	30	0.38	15.6%

F. Dimensionality Scaling

Optimization quality improves with morphology dimensionality, demonstrating scalability:

- 5 DOF (legs only): 8.1% CoT improvement
- 15 DOF (legs + torso): 12.3% CoT improvement
- 30 DOF (full body): 15.6% CoT improvement

G. Critical Design Parameters

Analysis reveals hip, knee, and ankle oblique angles account for 73% of walking CoT improvement, while upper body angles contribute primarily to cartwheel and jump kick performance—validating the biomechanical importance of lower-limb joint orientations.

V. DISCUSSION

A. Limitations

1) **Differentiable simulation requirement**: PGHC requires analytical gradients, limiting applicability to physics engines supporting automatic differentiation.
2) **Sim-to-real transfer**: Optimized morphologies are validated in simulation. Real-world deployment requires domain randomization and robustness analysis.
3) **Local optima**: Like all gradient-based methods, PGHC can converge to local optima. Random restarts or hybrid evolutionary-gradient approaches may improve global search.
4) **Manufacturing constraints**: Optimized parameters may not correspond to easily manufacturable configurations without additional constraints.

B. Future Work

1) **Hardware validation**: Real-world testing of optimized morphologies
2) **Multi-objective optimization**: Balance efficiency, robustness, and cost
3) **Expanded design spaces**: Joint optimization of geometry, actuators, and materials
4) **Hybrid evolutionary-gradient methods**: Combine global search with local refinement

VI. CONCLUSION

We present PGHC, a scalable framework for robot co-design that addresses the computational intractability of bilevel optimization through dynamic stability gating and adaptive trust regions. By rigorously enforcing the conditions under which gradient decoupling via the Envelope Theorem remains valid, PGHC enables efficient morphology optimization with deep RL policies. We demonstrate scalability on full-body humanoid morphology (30+ DOF), achieving 12-18% performance improvements across diverse locomotion tasks. Framework generality is validated through successful optimization of both joint orientations and link lengths, establishing PGHC as a practical approach for learning-based robot co-design.

REFERENCES