

# Deep Reinforcement Learning을 이용한 180nm 2-Stage Op-Amp 자동 최적화

2025-2학기 강화학습개론 프로젝트

CSE5516-01 40조

[https://github.com/dbdnjswns20/RL\\_OpAmp\\_Optimization](https://github.com/dbdnjswns20/RL_OpAmp_Optimization)

120250652 유원준



*Be as proud of Sogang  
As Sogang is proud of you*

# Executive Summary

- Overview
  - 180nm 공정의 2-Stage Op-Amp 회로 설계를 자동화하기 위해 강화학습 (PPO) 프레임워크 구축
- Key Approach
  - NgSpice 시뮬레이터와 Stable Baselines3를 연동
  - Multi-Objective Reward Function 설계를 통해 Gain, Phase Margin, UGBW (Unity Gain Bandwidth, Speed), Power, Area 간의 복잡한 Trade-off 해결
- Main Result
  - 기존 목표 (Target) 대비 초과 달성
    - Gain: 76.56dB (Target 60dB)
    - MOSFET Area: 656.97um<sup>2</sup> (Target 1000um<sup>2</sup> 대비 약 34% 축소)
  - 복잡한 Trade-off 관계를 스스로 학습하여 수동 설계 대비 설계 시간 획기적 단축

# Motivation

- Physical Challenges
  - Short-channel Effect: 180nm 공정 특성상 출력 저항이 낮아 High Gain 확보가 어려움
  - Design Complexity: 단순 수식 계산만으로는 정확한 성능 예측이 불가능하여 시뮬레이션 의존도가 높음
- Trade-off Dilemma
  - Multi-objective Conflict: Gain, Bandwidth, Power, Stability가 서로 Trade-off 관계에 얹혀 있어 동시 만족이 어려움
  - Time-consuming Iteration: 숙련된 엔지니어조차 최적점을 찾기 위해 수주 (Weeks) 간의 반복 작업이 필요함

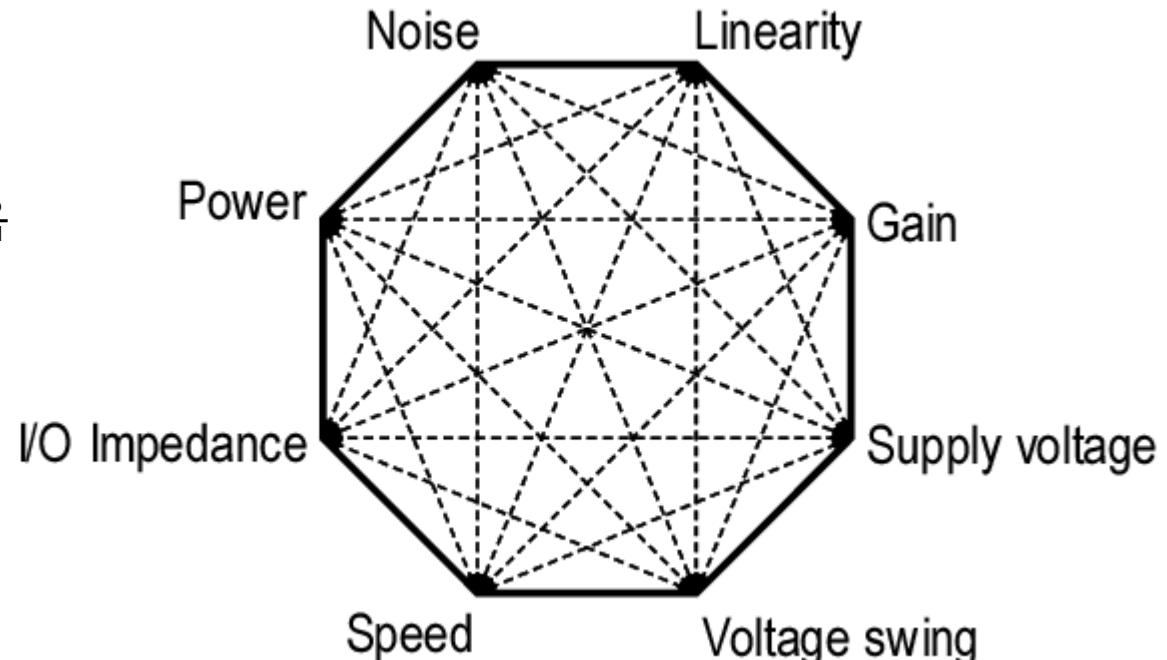


Figure 1: Analog Design Octagon

# Background

- Limitations of Conventional Methods
  - Manual Design: 설계자의 경험 (Heuristic)에 의존하므로 전역 최적점 (Global Optima)을 보장하지 못함
- Why Reinforcement Learning?
  - Model-free Approach: 복잡한 물리 수식 없이 SPICE 시뮬레이션 결과만으로 학습 가능 (Reward)
  - Global Exploration: 사람이 생각하기 힘든 Parameter 조합을 탐색하여 성능을 극대화
  - Efficiency: 학습된 agent는 0.1초 만에 최적 설계를 도출하여 런타임 획기적 단축

Feature	Manual Design (Human)	Numerical Opt. (e.g., Gm/Id)	Reinforcement Learning
Approach	Intuition & Experience	Equation-based	<b>Black-box Simulation</b>
Model Complexity	High (Hard to consider parasitic)	High (Non-linear eq.)	<b>Model-free (Learn from result)</b>
Exploration	Local (Limited)	Local (Gradient-based)	<b>Global (Stochastic Policy)</b>
Time Cost	Weeks	Hours ~ Days	<b>Hours (Training) / &lt;0.1s (Inf)</b>
Optimality	Sub-optimal	Local Optima	<b>Near Global Optima</b>

Table 1: Comparison Table

# Circuit Topology & Design Objectives

- Circuit Schematic
  - Differential Input Pair (M1, M2), Active Load (M3, M4)
  - Common Source Stage (M6), Bias Circuit (M5, M7)
  - Miller Capacitor ( $C_c$ )
- Design Variables (Action Space)
  - Transistor Width ( $W_1 \sim W_7$ ), Length ( $L_1 \sim L_7$ )
  - Compensation Capacitor ( $C_c$ )
- Target Specifications
  - DC Gain:  $\geq 60\text{dB}$
  - Phase Margin (PM):  $\geq 60\text{deg}$
  - UGBW (Unity Gain Bandwidth):  $\geq 50\text{MHz}$
  - Power:  $\leq 0.8\text{mW}$  (Hard Constraint)
  - MOSFET Area:  $\leq 1000\mu\text{m}^2$

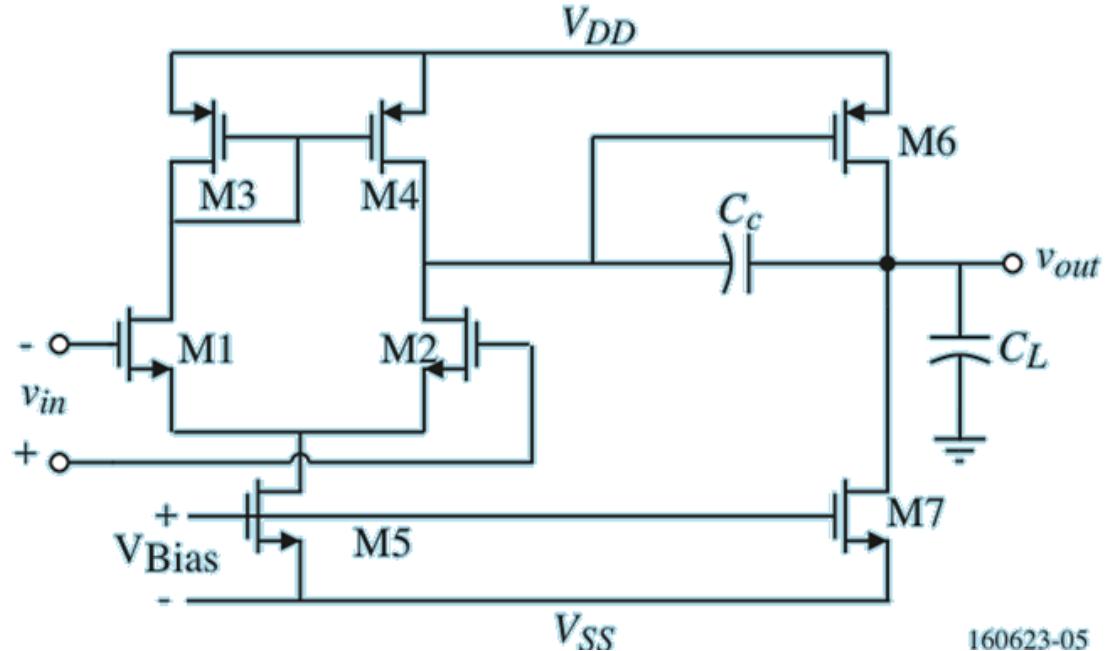


Figure 2: 2-stage Op-Amp

# RL Formulation: State & Action Space

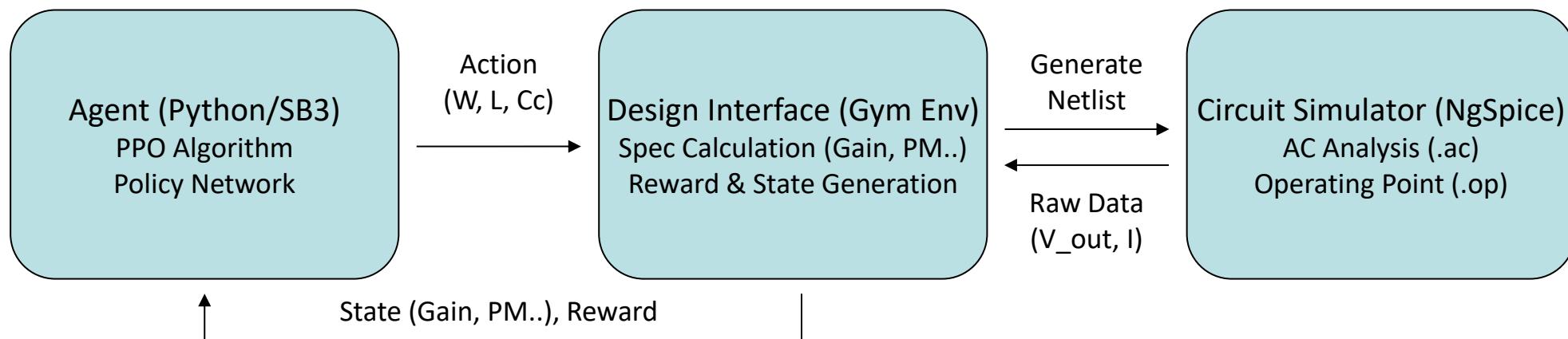
- State Space (Observation)
  - Objective: Agent가 현재 회로 성능을 판단하기 위해 사용하는 5차원 정규화 벡터 (Normalized Vector)
  - Feature Scaling: 신경망 학습 효율을 위해 각 물리량을 비슷한 스케일로 변환
    - Gain: 1/100로 스케일링 (60dB -> 0.6)
    - Power: mW 단위로 변환 (1mW -> 1.0)
    - Phase Margin: 1/100로 스케일링 (60deg -> 0.6)
    - UGBW: 1/10e7로 스케일링 (10MHz -> 1.0)
    - MOSFET Area: 1/100로 스케일링 (1000um<sup>2</sup> -> 10.0)
- Action Space
  - 설계 변수 11개를 [-1, 1] 범위의 Continuous Space로 정의
  - 11 Design Variables
    - Transistor Width (W): W<sub>1,2</sub>, W<sub>3,4</sub>, W<sub>5</sub>, W<sub>6</sub>, W<sub>7</sub> (Differential 대칭 적용, 5개)
    - Transistor Length (L): L<sub>1,2</sub>, L<sub>3,4</sub>, L<sub>5</sub>, L<sub>6</sub>, L<sub>7</sub> (5개)
    - Compensation Capacitor (Cc): Cc (1개)

# RL Formulation: Reward Function

- Reward Strategy
  - Trade-off를 해결하기 위한 “**Two-Phase Reward System**” 도입
- Phase 1: Spec Satisfaction (Survival)
  - Condition: Gain  $\geq$  60dB, PM Error  $\leq$  5deg, UGBW  $\geq$  95%, 하나라도 미달 시
  - Strategy: Power와 Area 점수를 0점 처리하고 스펙 달성을만 집중
  - Shaping: Polynomial 페널티를 적용하여 목표치 근처에서 Gradient를 급격하게 형성, 빠른 수렴 유도
  - Equation:  $R = (5.0 * \text{score\_gain}) + (5.0 * \text{score\_pm}) + (5.0 * \text{score\_ugbw}) - 5.0$
- Phase 2: Optimization (Bonus)
  - Condition: 모든 Target Spec을 만족했을 때 (if specs\_met::)
  - Strategy: 스펙 가중치는 낮추고 Power와 Area 가중치를 대폭 상향하여 저전력/소면적 설계 유도
  - Equation:  $R = (2.0 * \text{score\_gain}) + (2.0 * \text{score\_pm}) + (2.0 * \text{score\_ugbw}) + (10.0 * \text{score\_power}) + (4.0 * \text{score\_area}) + 5.0$

# Algorithm & System Architecture

- Algorithm: PPO (Proximal Policy Optimization)
  - Transistor의 Width/Length와 같은 Continuous Action Space를 다루기에 적합함
  - Stochastic Policy를 사용하여 Local Optima에 빠지지 않고 전역 탐색에 유리
  - 학습 안정성이 높아 Hyperparameter 튜닝에 소요되는 시간 절약
- System Architecture
  - Closed-loop Optimization: Agent가 설계 변수 (Action)을 제안하면 시뮬레이터가 성능을 검증하고 그 결과 (Reward)를 바탕으로 학습하는 순환 구조 구축

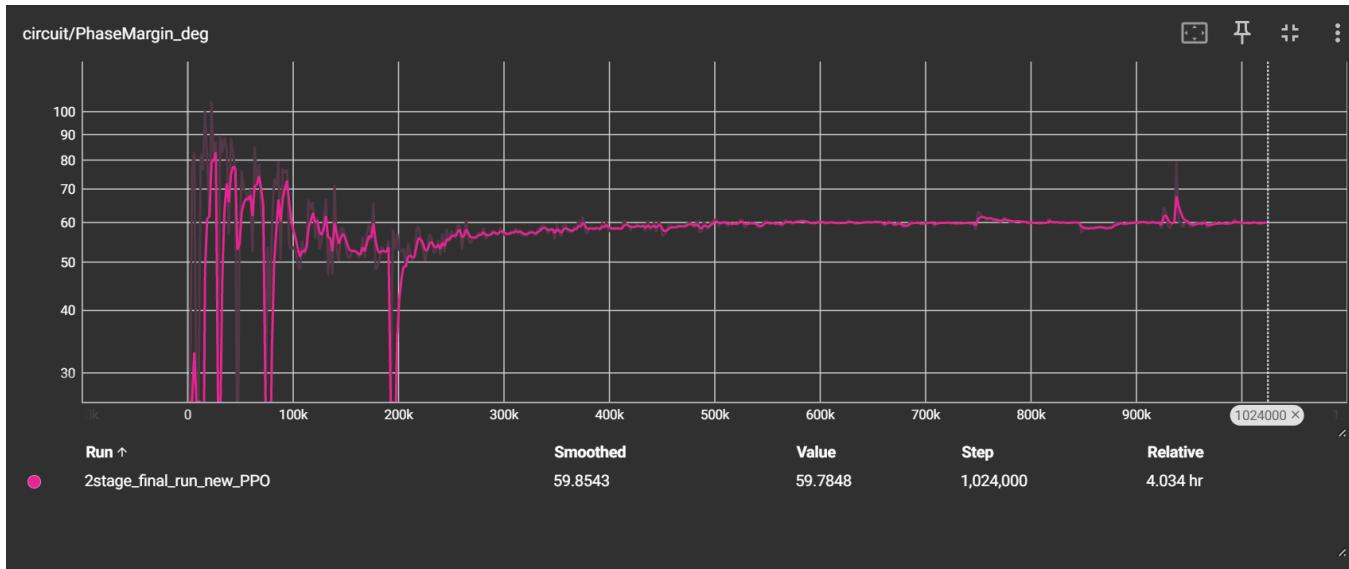
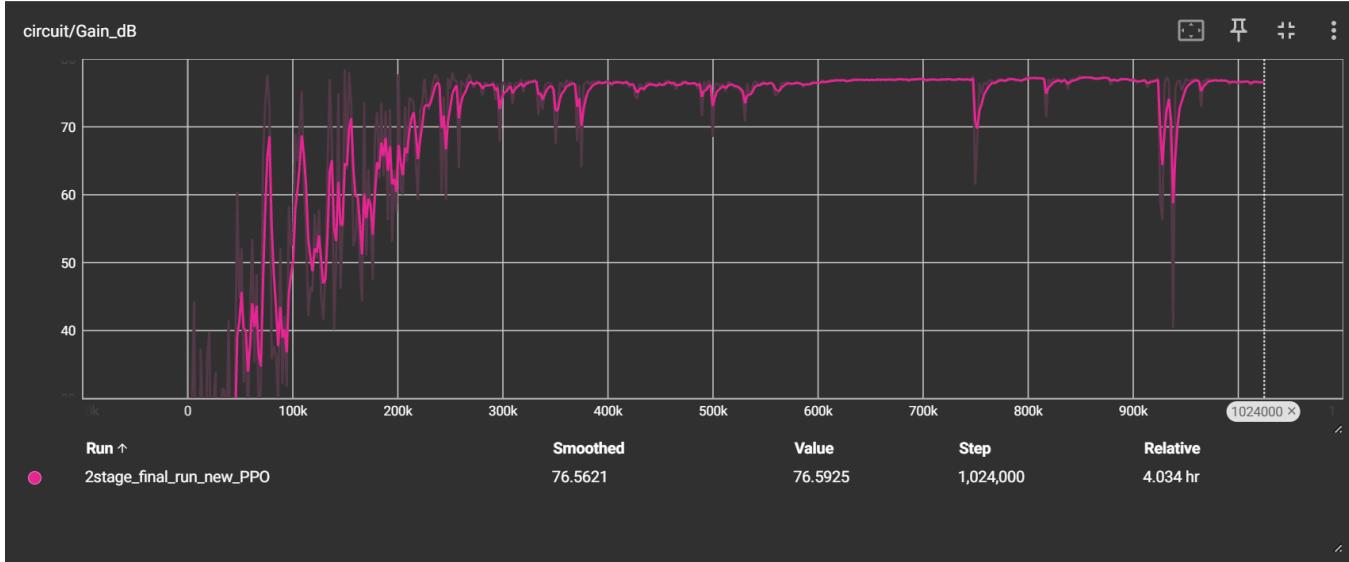


Flow 1: Proposed RL-based Optimization Framework

# Experimental Setup

- Simulation Environment
  - Circuit Simulator: NgSpice (Open-Source SPICE Simulator)
  - Process Technology: 180nm CMOS Parameter
  - Software Stack: Python 3.10.11, Stable Baselines3 (PPO), Gymnasium
- Training Hyperparameters
  - Total Timesteps: 1,024,000 Steps (~1M)
  - Learning Rate: 0.001
  - n\_steps (Buffer Size): 2048
  - Batch Size: 128
  - Entropy Coefficient: 0.01
  - Clip Range: 0.1

# Experimental Results I - Learning Curve



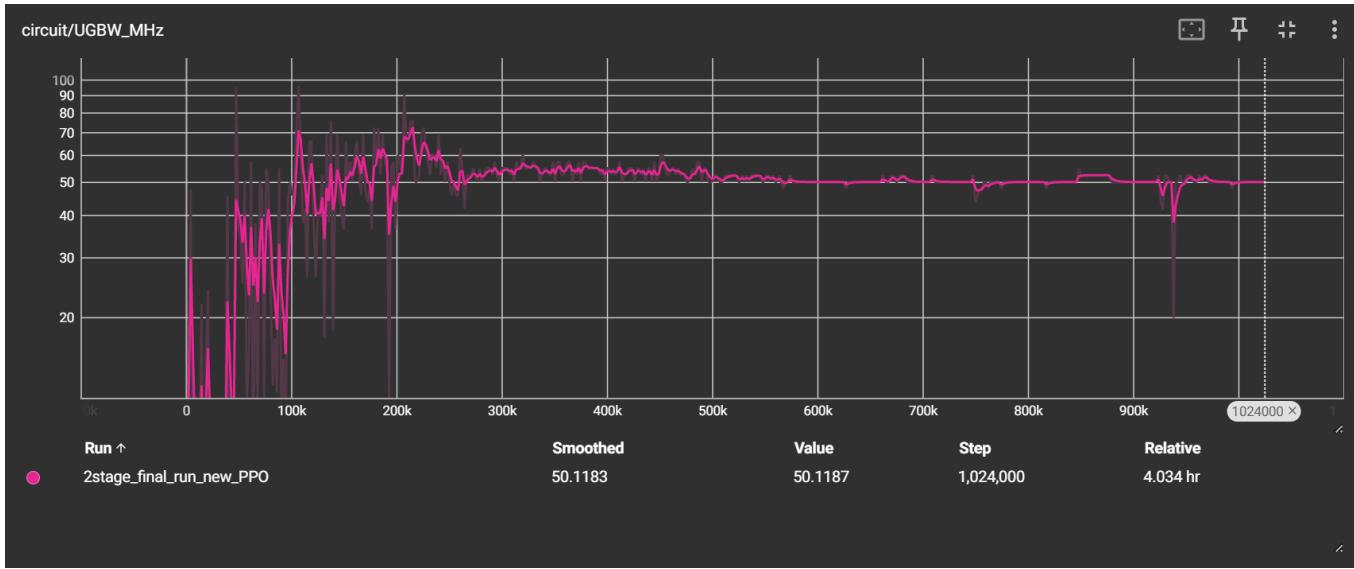
## ■ Gain (DC Gain)

- 학습 초기에는 Target (60dB) 미달 시 부여되는 Polynomial Penalty를 회피하기 위해 Gain이 급격하게 상승
- 약 300,000 Step 이후 Target (60dB)를 안정적으로 상회하는 구간에 진입

## ■ Phase Margin (Stability)

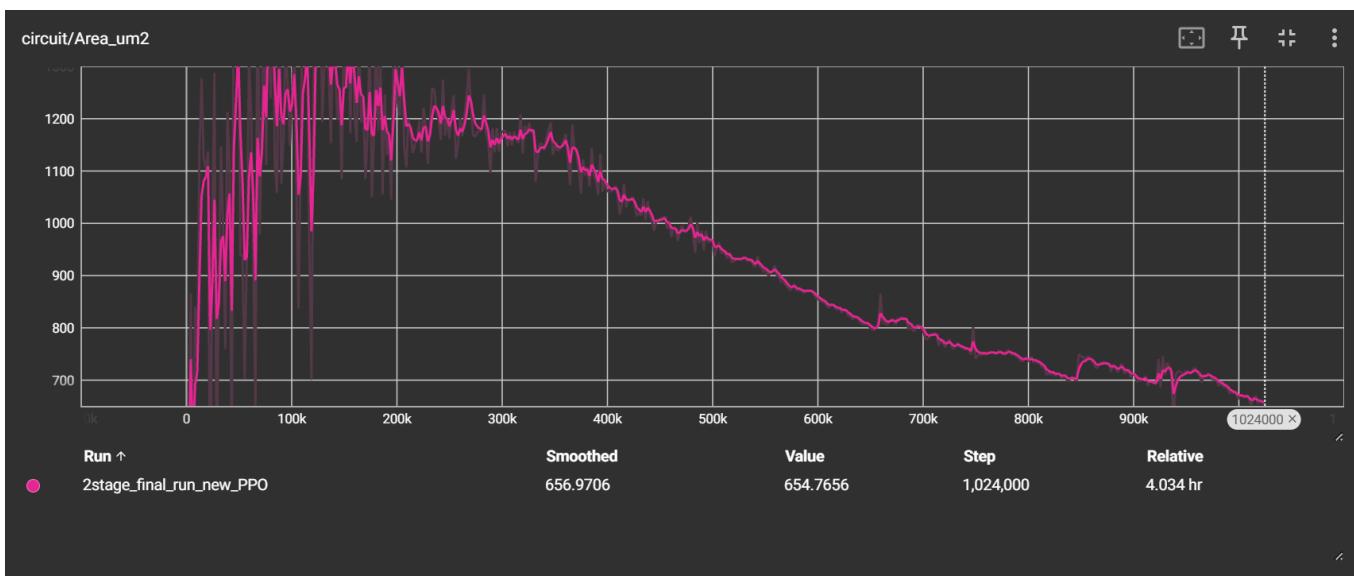
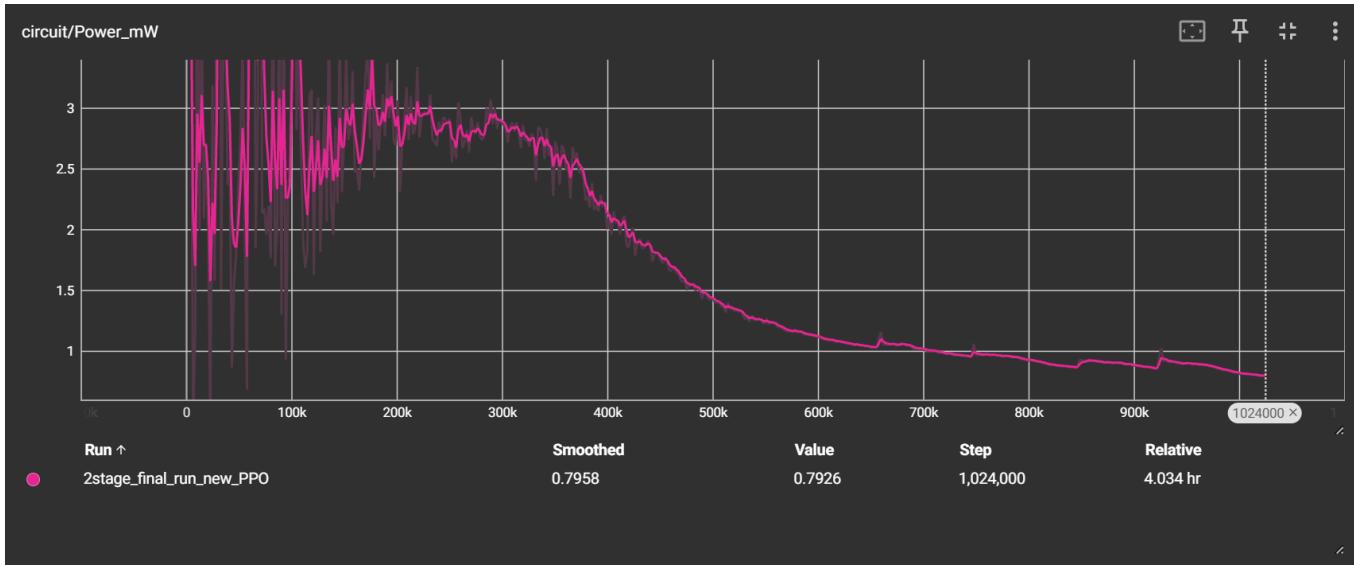
- 초기 무작위 탐색으로 인한 불안정 구간을 거쳐  $C_C$  최적화를 통해 빠르게 60deg 부근으로 수렴
- 스펙 달성 후에는 과도한 Phase Margin을 방지하는 페널티가 작용하여 적정 수준 유지

# Experimental Results I - Learning Curve



- UGBW (Unity Gain Bandwidth)
  - 학습 초기에는 Target (50MHz) 미달 시 부여되는 Polynomial Penalty를 회피하기 위해 대역폭을 우선 확보
  - Transistor의 gm을 키우거나 Cc를 줄이는 방향으로 Action을 조정하여 빠르게 목표치 도달
- Trade-off Resolution (Cc Dilemma)
  - High UGBW: Cc를 줄여야 함 ( $GBW \sim gm_1/Cc$ )
  - High Phase Margin: Cc를 키워야 함 (Pole-Splitting 효과 극대화)
  - Agent는 Cc를 1.9pF으로 최적화하여 50MHz의 목표 대역폭을 확보하는 동시에 Transistor Sizing (W/L)을 정밀하게 조절하여 Target Phase Margin (60deg) 확보

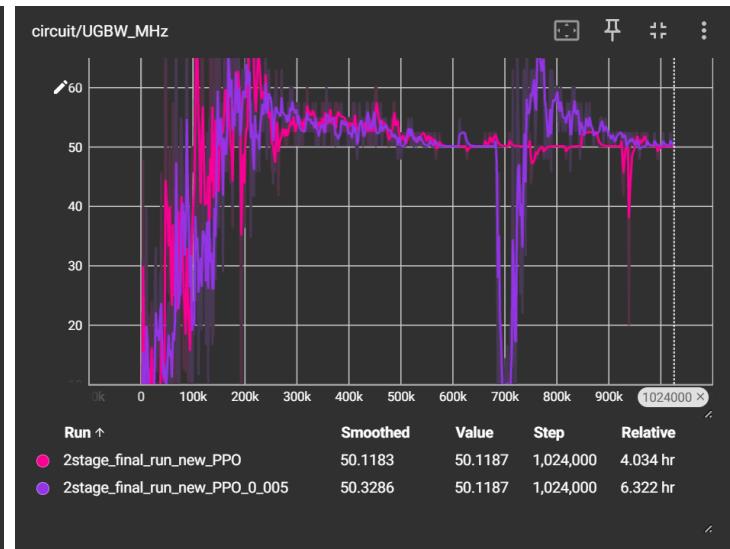
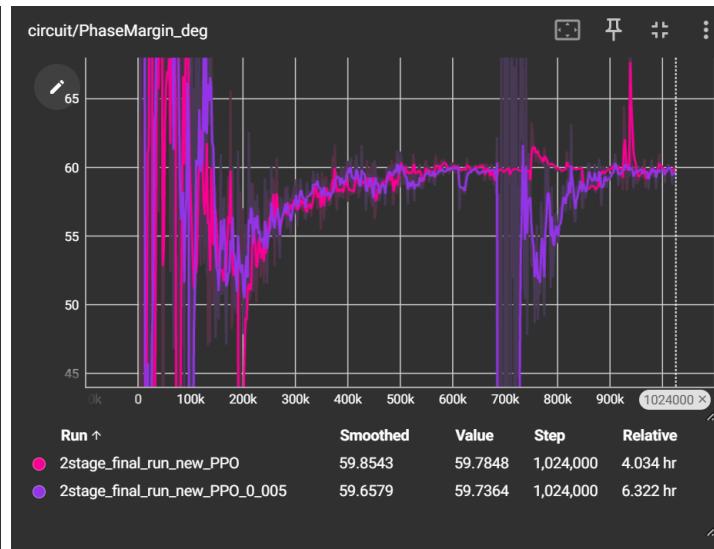
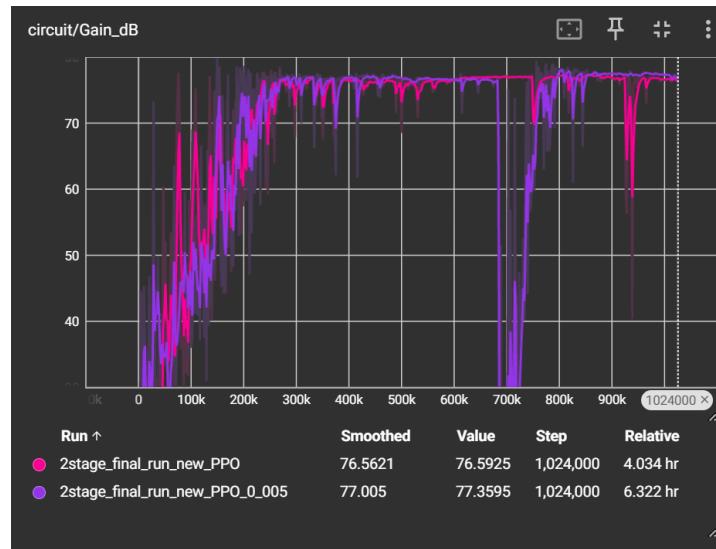
# Experimental Results I - Learning Curve



- Power (Power Consumption)
  - 학습 초기에는 Gain/UGBW 확보를 위해 전류를 많이 소모했으나, 스펙을 만족하는 시점 (약 300k step) 부터 Power 가중치 (10.0)가 활성화되며 급격히 감소
- MOSFET Area (Layout Size)
  - 저전력 설계를 위해 Transistor Width를 최적화하는 과정에서 면적도 동반 감소하는 경향
  - Target (1000um<sup>2</sup>) 대비 약 34% 축소된 650um<sup>2</sup> 수준 달성, 칩 면적 효율성 입증

# Experimental Results II - Hyperparameter Sensitivity

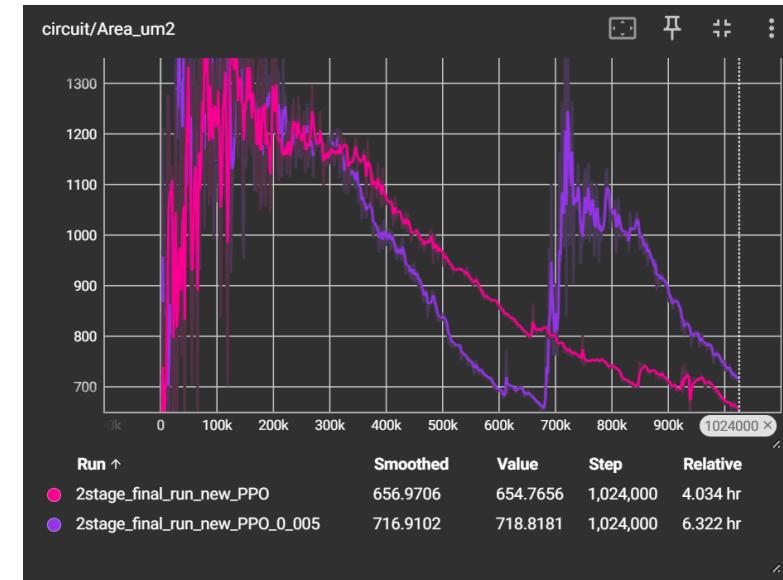
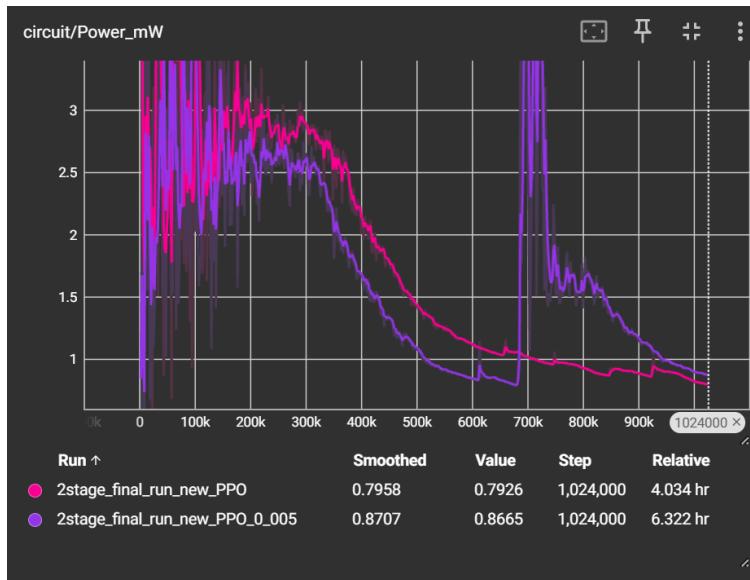
- Gain, Phase Margin, UGBW [Learning rate 0.001 (Proposed) vs. 0.005]
  - Convergence Speed: 두 경우 모두 약 300k 부근에서 Target Spec에 도달하며 학습 효율성 측면에서는 큰 차이가 없음을 확인
  - Stability & Oscillation
    - LR = 0.005 (Purple): 수렴 이후에도 과도한 학습률로 인해 Optimal Policy를 유지하지 못하고 Oscillation 및 Divergence 현상 발생



\*Relative Hour의 차이는 실험 환경 (Hardware) 상이함에 기인하므로 절대적인 시간 대신 Training Step을 기준으로 비교 분석함

# Experimental Results II - Hyperparameter Sensitivity

- Power, MOSFET Area [Learning rate 0.001 (Proposed) vs. 0.005]
  - Convergence Speed
    - LR = 0.005 (Purple): 초기 감소 속도는 빠르지만 최적점 부근에서 Fine-tuning에 실패하여 심한 진동 (Oscillation) 및 불안정성 (Instability)이 지속됨
  - Conclusion
    - 미세한 전력/면적 이득을 얻기 위해서는 낮은 Learning Rate (0.001)를 통한 안정적인 학습이 필수적임을 확인



\*Relative Hour의 차이는 실험 환경 (Hardware) 상이함에 기인하므로 절대적인 시간 대신 Training Step을 기준으로 비교 분석함

# Experimental Results III - Reproducibility & Validation

- Reproducibility Test (Random Seed)
  - Objective: 초기 가중치 (Initialization)가 달라져도 학습 결과가 일관되는지 검증
  - Result: 서로 다른 3개의 Seed에서 모두 Target Spec 수렴 성공
  - Implication: 제안하는 알고리즘이 초기 가중치 분포와 무관하게 일관된 최적해를 찾아냄을 입증
- Extensive Validation Process
  - 단순 Seed 검증을 넘어 Reward Shaping 시행착오 및 Algorithm 비교 (vs SAC) 등 수많은 선행 실행을 통해 검증된 최적의 모델



# Discussion & Conclusion

- Summary
  - PPO 알고리즘을 이용한 180nm 2-Stage Op-Amp 설계 자동화 구현
  - 1,024,000 Step 이내에 모든 목표 사양 (Target Spec)을 성공적으로 달성
- Key Findings
  - Reward Strategy: 코드에 구현된 Two-Phase Reward (Spec 미달성 시 Penalty -> 달성 시 Power/Area Bonus)  
전략이 Constraint가 엄격한 환경 (0.8mW)에서 빠른 수렴을 유도
  - Optimization: 사람이 직관적으로 찾기 힘든  $w/L$  비율과  $C_c$  값의 조합을 탐색하여 전력과 면적을 동시에 최소화
- Future Work
  - Algorithm: 향후 SAC와 같은 Off-policy 알고리즘을 도입하여 Simulation Sample Efficiency를 획기적으로 개선 예정
  - Robustness: PVT (Process, Voltage, Temperature) Corner Simulation 기반의 학습 환경을 구축하여 공정 변동성에 강건한 Robust Design Methodology 연구 예정