Priority-Based Graph-Enhanced Reinforcement Learning for Robust Analog Circuit Optimization

Jintao Li[1], Zhenxin Chen[1, 2], Sicheng He[1], AoJin Li[1], Shui Yu[1*]

[1]Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, China
[2] School of Electronics and Communication Engineering, Guangzhou University, Guangzhou 510006, China

Abstract

A primary motivation for analog integrated circuit (IC) design automation is the inefficiency of manual design in meeting increasingly stringent specifications, which often involve over 10 objectives. Recent advances in reinforcement learning (RL) emerge as a promising method, yet gaps remain when considering full design specifications, especially under process-voltage-temperature (PVT) variations. Excessive objectives lead to diminished reward signals, while varying PVT conditions result in conflicting gradients, both of which result in inefficient exploration. To address these, we propose a priority-based graph-enhanced RL framework. Specifically, using fuzzy logic converts quantitative rewards into qualitative priority signals, mitigating reward deterioration and enhancing exploration via entropy regularization. Furthermore, a graph-based representation compresses high-dimensional objective spaces under PVT variations into low-dimensional manifolds, enabling dynamic resource allocation to variation-sensitive regions and resolving gradient conflicts. Empirical results on various real-world analog ICs demonstrate that our method significantly outperforms existing RL algorithms, achieving superior solution quality and reducing simulation overhead.

Introduction

Mixed-signal integrated circuits are crucial in modern electronics, but while digital design benefits from automated tools, analog design still heavily relies on human expertise. The process is time-consuming and complex, beginning with topology analysis and performance equation derivation. This is followed by initial sizing, iterative simulations, and refinements to meet specifications (Chen et al. 2025). The large design space, extensive simulations, and performance tradeoffs contribute to significant effort, with additional challenges arising from performance deviations caused by process, voltage, and temperature (PVT) variations (Li et al. 2024; Cai et al. 2025). These challenges have driven research into automated analog sizing methods to improve efficiency and reduce manual input (Gielen, Walscharts, and Sansen 1990; Liu et al. 2021; Zhi et al. 2025b).

Traditional approaches typically treat analog sizing as a black-box optimization problem, utilizing machine learning (ML) techniques such as Bayesian optimization (BO) (Lyu et al. 2018; Gu et al. 2024b,a) and evolutionary algorithms (EA) (Li et al. 2023a; Budak et al. 2022; Li et al. 2025b) to navigate the expansive design space. These methods primarily focus on performing extensive computations to achieve the desired performance specifications, often sidestepping the circuit's intrinsic behavioral characteristics (Ding et al. 2023; Xu et al. 2024; Zhong et al. 2025). To address this limitation, recent efforts have incorporated reinforcement learning (RL) (Wang et al. 2018; Choi et al. 2023), which models the sizing problem as sequential decision-making. By learning adaptive policies through interactive exploration and feedback, RL efficiently navigates high-dimensional spaces and captures parameter-performance relationships beyond black-box approximations (Zhi et al. 2025a; Li et al. 2026). Furthermore, to mitigate the performance deviations induced by PVT variations, multi-task RL (MTRL) strategies (Shi et al. 2022; Kong et al. 2024) have emerged as a prominent solution. In this approach, PVT variations are treated as distinct optimization tasks, with transfer learning employed to expedite the optimization process and enhance the overall robustness of the design.

Despite its potential, applying RL to analog sizing presents significant challenges. **Reward Signal Deterioration:** When considering the full specifications, analog circuits often involve over ten constraints and objectives. As the policy improves, the advantage values shrink, leading to diminishing gradients. This causes gradient vanishing, slow convergence, and difficulty in navigating large search spaces during many-objective optimization (Pan et al. 2025). **Conflicting Gradients from PVT Variations:** PVT variations cause conflicting gradients across different conditions, leading to stagnation or oscillations in optimization. This disrupts RL's ability to converge effectively (Luo et al. 2023), as it relies on consistent feedback to update its policies (Li et al. 2023b).

To address the low exploration efficiency caused by PVT variations in the objective space, we represent the relationship between objectives and PVT variations using a graph-based approach. Unlike MTRL methods that treat each PVT scenario as an independent optimization task, our approach compresses the high-dimensional objective space by modeling the continuous Pareto manifold in a low-dimensional graph. This graph-based compression enables dynamic ac-

*Corresponding author. (yushui@uestc.edu.cn)

tivation of locally relevant regions of the objective space, allocates exploration resources to PVT-sensitive Pareto frontiers, and thereby significantly accelerates convergence. Furthermore, to address the reward degradation caused by many objective trade-offs, we convert quantitative reward signals into qualitative priority signals using fuzzy logic. This approach focuses on the relative superiority of objectives rather than their absolute reward values, reducing the dependency on reward magnitude and mitigating the diminishing return issue, thereby stabilizing the learning process. By extending our method with entropy regularization objectives, we facilitate efficient exploration in high-dimensional optimization spaces, especially under PVT variations.

- **Fuzzy Logic-based Reward Transformation:** We propose a method that converts quantitative rewards into qualitative priority representations through fuzzy logic, reducing dependence on reward magnitudes. This tackles the reward degradation issue arising from many-objective trade-offs in RL, mitigating diminishing returns and enhancing the consistency of policy convergence.
- **Entropy-Regularized Reward Reparameterization:** By reparameterizing rewards with entropy regularization and synergizing with fuzzy logic, this approach balances exploration and exploitation in RL. This ensures directional exploration aligns with fuzzy logic priorities, preserving solution diversity while accelerating convergence to robust many-objective policies.
- **PVT-Graph Enhanced RL Method:** We propose a PVT-aware graph representation method, which models the multi-task many-objective space into a low-dimensional graph. This approach effectively mitigates the dynamic reward distribution drift in RL caused by PVT variations, thereby resolving the challenges of policy update oscillations and value function convergence difficulties.

Experiments on AnalogGym (Li et al. 2025a), a real-world and open-source analog circuit test suite, demonstrate that our approach outperforms state-of-the-art methods by achieving faster convergence and higher solution quality.

Background

PVT Variations

PVT variations are inevitable phenomena in semiconductor design, systematically modeled through corners to ensure robust circuit performance in real-world scenarios (Jain et al. 2024). **Process corners** quantify manufacturing-induced variations in MOSFET, including slow-slow (SS), fast-fast (FF), slow-fast (SF), and fast-slow (FS); **voltage corners** (e.g., ±10% deviations from a 1.0V nominal value) reflect power supply fluctuations; and **temperature corners** (-40°C to 125°C, with 25°C as the nominal reference) capture environmental effects and circuit self-heating. These variations directly impact critical transistor performance metrics, such as speed and power consumption, causing deviations in overall circuit performance under different PVT conditions. Since worst-case PVT conditions shift with design parameters, the optimization process must ensure all performance metrics meet design specifications across dynamically changing PVT scenarios.

Related Work: Automatic Analog Sizing with RL

GNN-Enhanced Topology-Aware framework The rising popularity of RL in analog sizing is largely driven by GNN-enhanced frameworks, which address key limitations of early RL approaches by integrating domain knowledge of circuit topology (Wang et al. 2020). These frameworks enable ”open-box optimization” by modeling circuit topology graphs with GNNs, where node and edge embeddings encode component connections and constraints. By evolving beyond this basic encoding to enhance topological embeddings (Li and Carusone 2023; Hakhamaneshi et al. 2023), GNN-RL more effectively captures nuanced component interactions, thereby significantly boosting precision and interpretability. While most of these works have focused primarily on advancements in architecture or learning paradigms, less effort has been devoted to developing novel RL-based optimization frameworks.

Multi-Agent Coordination Optimization Framework Multi-agent RL (MA-RL) mirrors expert practice by splitting a complex circuit into subblocks, assigning one agent per subblock (Choi et al. 2024), and allowing agents to communicate so they can trade off local objectives while meeting global specs (Bao et al. 2024; Zhang et al. 2023). This decomposition turns a high-dimensional search into coordinated subproblems. However, the overhead of designing and training many agents often erodes the time savings, and local (subblock-level) errors can accumulate and propagate, making it hard to assess the true end-to-end impact of the RL policy.

PVT-Aware Robust Optimization Framework A critical focus is mitigating the susceptibility of analog circuits to PVT variation through strategies that balance awareness and efficiency. Frameworks like RobustAnalog (Shi et al. 2022) and PVTSizing (Kong et al. 2024) model PVT scenarios as related tasks, using task similarities to reduce objective competition. RobustAnalog prunes redundant tasks dynamically, whereas PVTSizing introduces random mismatches. Complementing this, hybrid approaches (Gao, Cao, and Zhang 2023; Cao et al. 2025) use BO to initialize RL with near-optimal datasets and enable batch-sampling parallelization.

While prior work often focuses on limited key specifications, our approach leverages fuzzy logic for reward-to-priority conversion, entropy regularization for exploration-exploitation balance, and a PVT-aware graph to compress high-dimensional spaces, effectively resolving reward degradation and PVT-induced gradient conflicts in full-spec analog sizing.

Problem Definition

We formulate robust analog sizing as a many-objective optimization under the PVT variations problem. Let x represent the continuous design parameter vector, encompassing transistor dimensions, passive components, and bias currents.

For each corner (P, V, T), circuit simulation yields a performance vector $\mathbf{F}(x, P, V, T)$. Our objective is to find an optimal design $x^\star$ that satisfies all design specifications while optimizing this k-dimensional overall performance under variations:

$$\begin{aligned} \min_{x} \ & \mathbf{F}(x, P, V, T) = [f_1(x, P, V, T), \ldots, f_k(x, P, V, T)] \\ \text{s.t.} \ & g_i(x, P, V, T) \leq 0, \ i = 1, \ldots, m, \\ & h_j(x, P, V, T) = 0, \ j = 1, \ldots, n, \\ & (P, V, T) \in \mathcal{C}_{\text{PVT}}, \end{aligned} \tag{1}$$

where $\mathcal{C}_{\text{PVT}}$ denotes the set of all selected corners (e.g., 5 process corners $\times$ 3 voltage corners$\times$ 3 temperature corners), $k \gg 3$ denotes the number of objectives.

Framework

In this section, we first outline why standard RL struggles to satisfy full, multi-spec design targets. We then introduce our method, which accelerates training by assigning relative priorities to generated solutions and solving them with a graph-enhanced module. Finally, we show how to plug this method into existing GNN-RL pipelines (see Fig. 1). For background, the baseline GNN-RL architecture follows (Li and Carusone 2023) and is not repeated here.

Priority-Based Method for Sizing Problem

RL trains agents to maximize cumulative rewards through interaction with the environment, guided by numerical reward signals. In analog sizing under PVT variations problems, where state transitions exhibit stochasticity and delayed feedback, policy networks are commonly employed to output continuous action distributions. A prevalent approach (Sutton, Barto et al. 1998) is to update the policy parameters θ according to the following gradient estimator:

$$\nabla_\theta J(\theta) \approx \frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}} \frac{1}{|S_x|} \sum_{\tau \in S_x} \left[(r(x, \tau) - b(x)) \nabla_\theta \log \pi_\theta(\tau | x) \right], \tag{2}$$

Here, $\mathcal{D}$ denotes the dataset of problem instances; S_x is the set of sampled solutions for instance x, where each trajectory $\tau = (a_1, a_2, \ldots, a_T)$ comprises a sequence of continuous circuit parameter actions. $r(x, \tau)$ is the reward function derived from parameter optimization, and $b(x)$ serves as a baseline to compute the advantage $A(x, \tau) = r(x, \tau) - b(x)$, reducing gradient estimate variance. The policy $\pi_\theta(\tau|x)$ defines the trajectory distribution given instance x; in continuous action spaces, it is typically modeled as a Gaussian distribution $\pi_\theta(a_t|s_{t-1}) = \mathcal{N}(\mu_\theta(s_{t-1}), \sigma_\theta(s_{t-1}))$, leading to $\pi_\theta(\tau|x) = \prod_{t=1}^{T} \pi_\theta(a_t|s_{t-1})$.

For full-specification scenarios, baseline selection poses a critical challenge. In multi-objective settings, inadequate baselines amplify statistical errors in baseline estimation. The variance of the advantage function satisfies $\text{Var}[A(x, \tau)] \geq \sum_{k=1}^{K} \text{Var}[r_k(x, \tau)]$, with the lower bound growing linearly with the number of objectives K, destabilizing training. Additionally, as the policy improves, advantage magnitudes diminish across objectives, causing $A(x, \tau)$ to approach zero. This renders updates to the policy objective $J(\theta)$ negligible, trapping optimization in local minima.

Fuzzy Many-Objective Priority-Based Method

We propose a priority-based method that integrates fuzzy logic with many-objective optimization. This framework maps multi-dimensional quantitative rewards to qualitative priorities via fuzzy membership functions and resolves inter-objective conflicts through a dynamic priority aggregation operator (DPAO). For K objective functions $\{r_k(x, \tau)\}_{k=1}^{K}$, each objective is assigned a fuzzy priority $\mu_k \in [0, 1]$, with the priority label generation mechanism structured as follows.

The first core component is the fuzzy priority transformer. For each objective k, a triangular fuzzy number (a_k, b_k, c_k) defines its "high-priority" interval, and a membership function implements the quantitative-to-qualitative mapping:

$$\mu_k(r_k) = \begin{cases} 0 & r_k \leq a_k \\ \frac{r_k - a_k}{b_k - a_k} & a_k < r_k \leq b_k \\ \frac{c_k - r_k}{c_k - b_k} & b_k < r_k \leq c_k \\ 0 & r_k > c_k \end{cases} \tag{3}$$

Objective k exhibits non-zero priority only when $r_k \in [a_k, c_k]$, with the maximum priority achieved at $r_k = b_k$.

The second key component is the DPAO, which aggregates fuzzy priorities into a unified global priority strength. Using a learnable weight vector $\mathbf{w}(x) \in \Delta^{K-1}$ (where Δ^{K-1} denotes the $(K-1)$-dimensional simplex), the global priority strength is computed as:

$$\Psi(\tau|x) = \bigoplus_{k=1}^{K} w_k(x) \odot \mu_k(r_k) = \frac{\sum_{k=1}^{K} w_k(x) \cdot \mu_k^{p(x)}}{\sum_{k=1}^{K} \mu_k^{p(x)}} \tag{4}$$

Here, $p(x) \geq 1$ is a dynamic power exponent controlling aggregation "sharpness": as $p(x) \to \infty$, the DPAO simplifies to selecting the maximum priority among objectives $(\max_k \mu_k(r_k))$.

For generating priority labels between trajectory pairs (τ_1, τ_2), we define a threshold-based priority relation $\succ_\epsilon$, where $\Psi(\tau_1|x) \succ_\epsilon \Psi(\tau_2|x)$ if and only if $\Psi_1 > \Psi_2 + \epsilon$. The priority label y is then given by:

$$y = \mathbf{1} \left(\Psi(\tau_1|x) \succ_\epsilon \Psi(\tau_2|x) \right) \tag{5}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, which takes the value 1 if the condition inside the parentheses is satisfied and 0 otherwise. This threshold ϵ mitigates label fluctuations caused by trivial differences in priority strengths, thereby enhancing the consistency of learned priorities.

Fuzzy Priority Policy Optimization under the Maximum Entropy

In terms of policy optimization, we ground the framework in the maximum entropy framework, where the optimal policy must satisfy both fuzzy priority constraints and multi-objective tradeoffs. Formally, the optimal policy is proportional to:

$$\pi^*(\tau|x) \propto \exp\left(\alpha^{-1} \sum_{k=1}^{K} \lambda_k \mu_k(r_k) \right), \text{ s. t. } \lambda_k = f_k(\nabla \Psi), \tag{6}$$

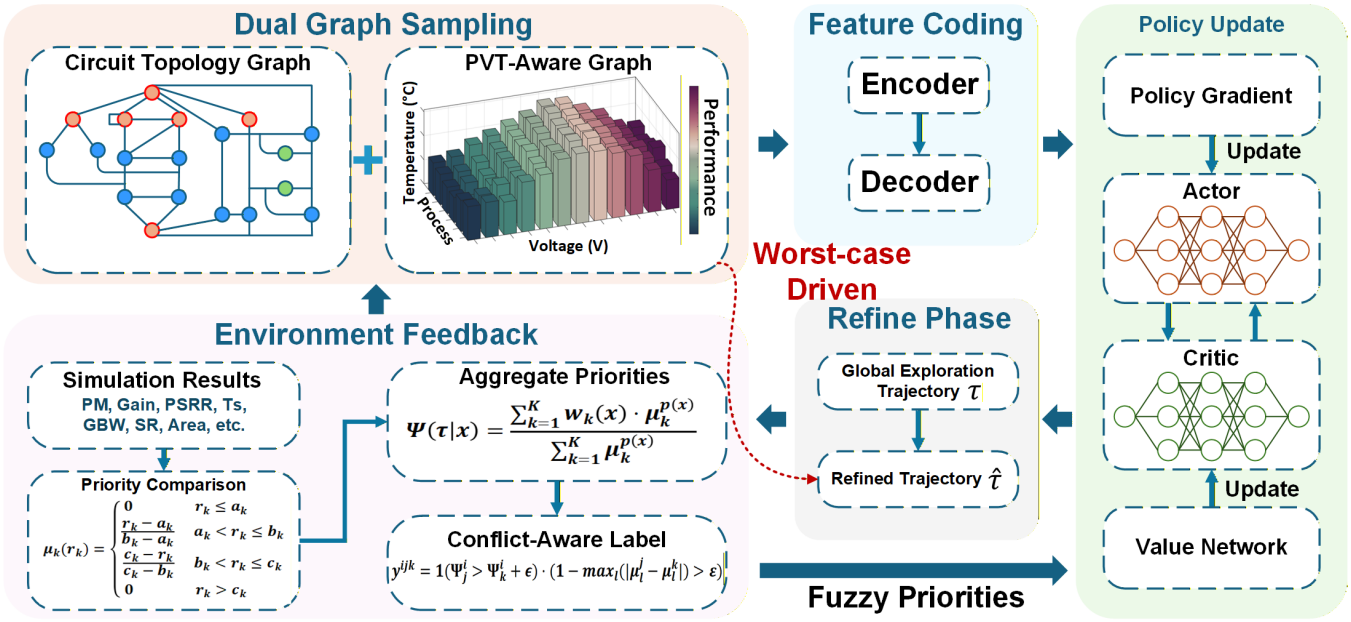

Figure 1: Algorithmic framework of fuzzy-priority graph-enhanced RL for analog circuits. In the priority aggregation, rewards are mapped to fuzzy priorities. The PVT-aware refinement refines trajectory $\boldsymbol{\tau}$ into $\hat{\boldsymbol{\tau}}$ under the worst-case driven.

Here, λ_k denote fuzzy Lagrange multipliers, which are dynamically adjusted via backpropagation of gradients from the DPAO to balance conflicting objectives.

Proposition 1 (Fuzzy Policy Equivalence) *Let* $\{\mu_k\}_{k=1}^K$ *be a set of complete trajectory features, and let the fuzzy priority space* Ψ *be defined as the quotient space of trajectory distributions under the equivalence relation induced by these features. If two policies* π_θ *and* π_ϕ *satisfy:*

$$\sum_{k=1}^{K} \lambda_k \left(\mathbb{E}_{\pi_\theta}[\mu_k(x)] - \mathbb{E}_{\pi_\phi}[\mu_k(x)] \right) = 0 \quad \forall \lambda \in \mathbb{R}^K, \forall x \in \mathcal{D} \tag{7}$$

where $\mathcal{D}$ *denotes the input space and* $\mathbb{E}_\pi[\mu_k(x)]$ *represents the expectation of feature* μ_k *over trajectories induced by policy* π *starting from* x*, then they induce equivalent trajectory distributions in the fuzzy priority space* Ψ*.*

The gradient update rule for the policy is derived using a fuzzy advantage function, defined as $\tilde{A}(\tau_1, \tau_2|x) = \Psi(\tau_1|x) - \mathbb{E}_{\tau'}[\Psi(\tau'|x)]$. The policy gradient is then formulated as:

$$\nabla_\theta J(\theta) = \mathbb{E}_{x,\tau_1,\tau_2} \left[\sigma \left(\tilde{A}(\tau_1, \tau_2|x) \right) \cdot \nabla_\theta \log \frac{\pi_\theta(\tau_1|x)}{\pi_\theta(\tau_2|x)} \right] \tag{8}$$

where $\sigma(\cdot)$ denotes the sigmoid function, introduced to smooth intransitive priorities and stabilize gradient updates.

Key properties of the proposed framework include fuzzy invariance and conflict resolution. Fuzzy invariance ensures that affine transformations of the reward functions r_k do not alter the relative order of fuzzy priorities μ_k, granting the policy gradient scale robustness. For conflict resolution, when there exist objectives i and j such that an increase in μ_i implies a decrease in μ_j (i.e., $\mu_i \uparrow \Rightarrow \mu_j \downarrow$), the DPAO automatically adjusts objective weights via $\mathbf{w}(x)$, steering the policy gradient $\nabla_\theta J(\theta)$ toward Pareto improvement directions.

PVT Variation Graph Representation

To precisely model PVT-dependent circuit behaviors, we construct a variation-aware graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$ that encodes the relationships between PVT variations and circuit performance. Formally, the graph is defined as:

$$\begin{cases} \mathcal{V} & = \{v_k \mid k \in [1, K]\}, \\ \mathcal{E} & = \{e_{ij} = \text{sim}(\mathbf{c}_i, \mathbf{c}_j) \mid i, j \in \mathcal{V}\}, \\ \mathcal{X} & = \{(\mathbf{c}_v, \mathbf{s}_v) \mid v \in \mathcal{V}\}, \end{cases} \tag{9}$$

where $\mathcal{V}$ denotes the set of PVT corners, $\mathcal{E}$ represents edges that quantify the similarity between corners through the function $\text{sim}(\mathbf{c}_i, \mathbf{c}_j)$, and $\mathcal{X}$ consists of node features integrating PVT parameters $\mathbf{c}_v$ and circuit performance metrics $\mathbf{s}_v$ for each node v in $\mathcal{V}$.

Each node v in the graph is characterized by a dual-stream feature vector $\mathbf{x}_v$, which integrates PVT parameters and circuit performance metrics to enable fine-grained modeling of operating conditions. This vector is defined as:

$$\mathbf{x}_v = \underbrace{[P^{(1)}, \ldots, P^{(m)}, V, T]}_{\mathbf{c}_v \in \mathbb{R}^{m+2}} \oplus \underbrace{[s_1, \ldots, s_n]}_{\mathbf{s}_v \in \mathbb{R}^n} \tag{10}$$

Here, $\oplus$ denotes feature concatenation. The PVT parameter stream $\mathbf{c}_v$ includes process variations encoded via one-hot vectors ($P^{(i)} = \mathbb{I}_{\{\text{process}=i\}}$, where $\mathbb{I}(\cdot)$ is the indicator function) and voltage/temperature values normalized to $[0, 1]$ within their operating ranges. The performance stream

$\mathbf{s}_v$ captures the circuit performance (e.g., gain, bandwidth) measured at corner v, linking PVT conditions to functional behavior.

The graph's edges are defined by a physically grounded adjacency matrix $\mathbf{A} \in \mathbb{R}^{K\times K}$, which quantifies the similarity between operating corners based on their PVT parameters. The matrix entries are given by:

$$\mathbf{A}_{ij} = \exp\left(-\gamma \left\|\mathbf{c}_i - \mathbf{c}_j\right\|^2_{\Sigma^{-1}}\right) \cdot \mathbb{I}_{\{\|\Delta \mathbf{c}_{ij}\|_2 < \tau\}} \tag{11}$$

This formulation combines a Gaussian kernel with a locality constraint: the Mahalanobis distance $\|\cdot\|_{\Sigma^{-1}}$ captures inherent dependencies between PVT variables, while γ controls the sensitivity to parameter differences. The indicator function $\mathbb{I}(\cdot)$ enforces sparsity by limiting edges to corners with small PVT deviations ($\Delta \mathbf{c}_{ij} = \mathbf{c}_i - \mathbf{c}_j$) via threshold τ, ensuring physical plausibility.

For predicting performance at new or unseen PVT corners v^*, the graph leverages a propagation mechanism via GNNs. The predicted performance $\hat{\mathbf{s}}_{v^*}$ is computed using information from the corner's local neighborhood:

$$\hat{\mathbf{s}}_{v^*} = \text{GNN}_\theta\left(\mathbf{A}, \{\mathbf{x}_v\}_{v\in\mathcal{N}(v^*)}\right), \tag{12}$$

Here, $\mathcal{N}(v^*)$ denotes the neighborhood of v^*, consisting of its k-nearest neighbors in the PVT parameter space with $k = O(\log K)$. This design balances efficiency and accuracy by focusing on physically relevant local interactions rather than global graph structure.

The PVT-graph aligns with physical and statistical principles to deliver key practical benefits. Dual-stream encoding separates PVT parameters ($\mathbf{c}_v$) and performance metrics ($\mathbf{s}_v$) into distinct spaces, thereby decoupling environmental inputs from circuit responses and enhancing the interpretability of PVT-performance relationships. Variation-aware edge weights $\mathbf{A}_{ij}$ encode physical regularity by quantifying the conditional correlation probability $p(\mathbf{s}_j \mid \mathbf{s}_i, \Delta\mathbf{c}_{ij})$, ensuring similarity metrics reflect real-world PVT dependencies. The locality threshold τ induces sparse connectivity ($|\mathcal{E}| = O(K \log K)$), mirroring the exponential decay of PVT impact in practice while reducing computational complexity.

Framework Overview

To address challenges in quantifying and balancing multiple-objective rewards and gradient conflicts resulting from insufficient policy robustness under PVT variations, the key idea of our framework is joint modeling of topology and environment, as well as reward-priority conversion, as shown in Algorithm 1. Our framework employs dual graph sampling. The circuit topology graph encodes component connectivity and constraints, forming the RL exploration space. The PVT graph models variation distributions and correlations, evaluating performance under uncertainties. Sampling diverse G_p instances ensures the policy learns PVT-robust solutions. Subsequently, the policy network π_θ first generates parameter trajectories τ using G_t for global exploration to avoid local optima. During late training, it refines τ into $\hat{\tau}$ by focusing on worst-case scenarios from G_p^i. This addresses the limitation of global exploration, missing extreme PVT variations, and ensuring robustness through targeted worst-case refinement.

Algorithm 1: Fuzzy Priority-Based Graph-Enhanced RL

Input: Circuit topology graph G_t; PVT graph G_p; Training steps T; TFT: Fine-tuning steps; Reward functions $r = [r_1, \ldots, r_k]$; Fuzzy parameters $\{a_k, b_k, c_k\}$; Aggregation parameters $\{w_k, \varepsilon\}$; Initialized policy network π_θ; Initialized value network V_ω

for $step = 1$ **to** $T + TFT$ **do**
  Sampling $G_t^i \sim D_t$, $G_p^i \sim D_p$
  Action Generation: $\tau^i \leftarrow \pi_\theta(G_t^i)$
  if $step > T$ **then**
    $\hat{\tau}^i \leftarrow refine\_parameters(\tau^i, r, G_p^i)$
    $\tau^i \leftarrow \tau^i \cup \hat{\tau}^i$
  end if
  $\mu_k(r_k), \Psi_j^i \leftarrow$ Fuzzy priority calculation according to Eq.(3) and Eq.(4).
  Conflict-Aware Label Generation:

$$y^{ijk} = \mathbf{1}(\Psi_j^i > \Psi_k^i + \epsilon) \cdot (1 - \max_l(|\mu_l^j - \mu_l^k|) > \varepsilon)$$

  Policy Gradient Update:

$$\nabla_\theta J(\theta) \propto \sum_{i,j,k} \sigma(\Psi_j^i - \mathbb{E}[\Psi]) \cdot y^{ijk} \cdot \nabla_\theta \log \frac{\pi_\theta(\tau_j^i|G_t^i)}{\pi_\theta(\tau_k^i|G_t^i)}$$

  Value Network Update:

$$\nabla_\omega J(\omega) \propto \sum_{i,j} (\Psi_j^i - V_\omega(G_p^i, \tau_j^i)) \cdot \nabla_\omega V_\omega(G_p^i, \tau_j^i)$$

  Parameter Updates:

$$\theta \leftarrow \theta + \eta \cdot \nabla_\theta J(\theta), \omega \leftarrow \omega + \eta \cdot \nabla_\omega J(\omega)$$

  $\{a_k, b_k, c_k\}, w_k \leftarrow$ Update fuzzy parameters and weights every 100 steps.
end for
return π_θ, V_ω

The fuzzy priority calculation is central to resolving multi-objective conflicts. The framework maps each objective reward to a [0,1] priority via the triangular fuzzy function $\mu_k(r_k)$, replacing numerical comparisons with priority judgments to eliminate scale differences. The DPAO then computes Ψ_j^i using learnable weights w_k to balance conflicting objectives automatically. Furthermore, conflict-aware label generation and gradient updates further boost learning stability. The labels filter two types of noise: minor priority differences to avoid resource waste on trivial distinctions, and high-conflict scenarios to pause learning during severe objective conflicts, thereby preventing misleading updates from erroneous signals. Policy gradients enhance better trajectory probabilities through contrastive learning, outperforming traditional single-trajectory reward updates, particularly in many-objective settings.

Algorithm		NMCF	NMCNR	ACBC	AFFC	DFCFC2	PFC	RAFFC
This work	$\text{FOM}_{AMP}\uparrow$	**754.3**	**523.6**	**274.8**	**41.5**	**7210.5**	**1964.7**	**65.2**
	Step $\downarrow$	**5917.6**	6215.4	5416.3	**5876.8**	**6524.2**	6109.5	5106.7
	Violations $\downarrow$	**3.2**	**1.5**	**2.1**	**3.7**	4.3	**2.8**	5.4
rGNN-RL (Li and Carusone 2023)	$\text{FOM}_{AMP}\uparrow$	382.5	281.7	152.6	18.9	3502.7	1002.4	32.8
	Step $\downarrow$	7215.8	7803.2	6805.7	8572.6	9204.5	8223.8	6503.2
	Violations $\downarrow$	18.7	15.9	16.3	20.6	25.3	20.1	22.9
MA-OPT (Choi et al. 2024)	$\text{FOM}_{AMP}\uparrow$	253.9	182.3	101.8	10.7	2203.5	701.6	19.5
	Step $\downarrow$	7502.3	11204.5	8803.6	11505.9	9302.8	11204.7	10207.5
	Violations $\downarrow$	30.5	25.8	28.4	35.2	40.7	32.5	38.6
PVTsizing (Kong et al. 2024)	$\text{FOM}_{AMP}\uparrow$	652.7	451.9	241.3	35.7	6003.2	1703.5	57.6
	Step $\downarrow$	6150.2	6783.5	5690.3	6354.7	7490.6	**5947.4**	**4981.6**
	Violations $\downarrow$	5.1	3.4	4.6	6.3	**3.1**	5.8	**4.2**
Rose (Cao et al. 2025)	$\text{FOM}_{AMP}\uparrow$	521.8	383.2	202.5	28.4	5001.8	1402.9	48.3
	Step $\downarrow$	6985.4	**6190.7**	**5109.9**	6861.3	6710.9	6291.2	5090.8
	Violations $\downarrow$	8.3	6.7	7.2	7.5	12.4	9.6	13.7

Table 1: Experimental results on seven complex circuits under PVT variations. FOM_{AMP} reflects worst-case performance under PVT variations. All data presented are mean values, and optimal results are highlighted in **bold**.

Experiment

In this section, we present key experimental results to validate the superiority of our proposed framework. We focus on addressing two critical questions: *1. How does the priority-based method perform against existing algorithms under full design specifications? 2. Does the PVT graph enable a better exploitation-exploration balance compared to MTRL?*

Testing Suite Setups

Performance evaluation of the proposed framework is conducted using the open-source test suite AnalogGym (Li et al. 2025a). Seven distinct complex circuits are selected for optimization, each involving 11 objectives, 5 constraints, and 20 PVT corners ($\{TT, FF, SS, FS, SF\} \times \{1.08V \sim 1.32V\} \times \{-25°\text{C} \sim 125°\text{C}\}$). Evaluations adhere to full design specifications, encompassing AC, DC, and transient analyses. For generality and fair comparison of many-objective optimization performance across algorithms, all specifications are treated equally. All simulations are executed in NGspice with the SKY130 PDK (Google and Foundry 2020), running on a workstation equipped with an Intel Xeon CPU, 64 GB memory, and an NVIDIA RTX 3080 GPU.

For consistent comparison across circuit instances under full design specifications, we adopt the AnalogGym-provided figure of merit, FOM_{AMP} (Li et al. 2025a), defined as:

$$\text{FOM}_{\text{AMP}} = \left(\frac{\text{PSRR}}{\text{PSRR}_{\text{ref}}} \cdot \frac{\text{CMRR}}{\text{CMRR}_{\text{ref}}} \cdot \frac{\text{Gain}}{\text{Gain}_{\text{ref}}} \cdot \frac{\text{FOM}_S}{\text{FOM}_{S,\text{ref}}} \cdot \frac{\text{FOM}_L}{\text{FOM}_{L,\text{ref}}}\right) \times \left(\frac{T_s}{T_{s,\text{ref}}} \cdot \frac{\text{Area}}{\text{Area}_{\text{ref}}}\right)^{-1} \text{FOM}_{\text{Penalty}}. \tag{13}$$

Here, PSRR denotes the power supply rejection ratio, CMRR the common-mode rejection ratio, and Gain the open-loop gain. FOM_{S} and FOM_{L} represent small-signal and large-signal performance metrics, while T_s refers to settling time and Area to circuit area. $\text{FOM}_{\text{Penalty}}$ is a penalty factor applied for constraint violations, and the subscript "ref" indicates reference values derived from baseline designs in AnalogGym. The FOM_S and FOM_L are defined as:

$$\begin{aligned} \text{FOM}_S &= \frac{\text{GBW} \cdot C_{\text{Load}}}{\text{Power}}, \\ \text{FOM}_L &= \frac{\text{SR} \cdot C_{\text{Load}}}{\text{Power}} \end{aligned} \tag{14}$$

In these expressions, GBW stands for the gain-bandwidth product, SR the slew rate, C_{Load} the capacitive load, and Power the total power consumption of the circuit. To penalize constraint violation, we define a multiplicative penalty factor:

$$\text{FOM}_{\text{Penalty}} = \left(\max\left(1, \frac{v_n}{v_{n,\text{ref}}}\right) \cdot \max\left(1, \frac{\text{TC}}{\text{TC}_{\text{ref}}}\right) \cdot \max\left(1, \frac{v_{\text{os}}}{v_{\text{os,ref}}}\right)\right)^{-1}, \tag{15}$$

Here, v_n is the output noise, TC is the temperature coefficient, and v_{os} is the input offset voltage. **Notably, FOM_{AMP} is used exclusively as a metric for comparing solution quality across algorithms in our experiments and does not participate in the optimization process.**

Baseline

We compare our framework against four baseline methods spanning distinct paradigms: rGNN-RL (Li and Carusone 2023), a GNN-enhanced framework; MA-OPT (Choi et al. 2024), a multi-agent optimization framework; Rose (Cao et al. 2025), a hybrid architecture-based method; and PVTsizing (Kong et al. 2024), an MTRL approach. Both rGNN-RL and MA-OPT enhance PVT robustness through worst-case-driven strategies (Li et al. 2024).

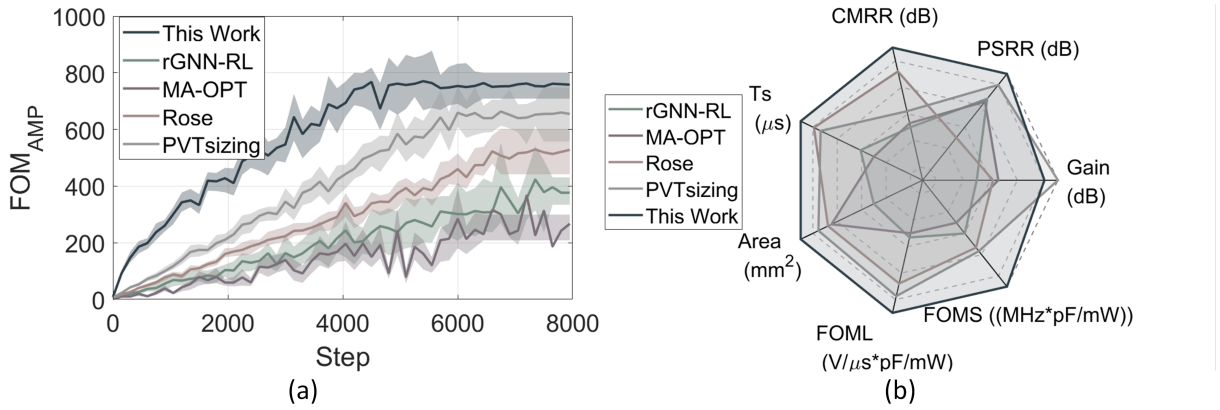

Figure 2: (a)-(b): Comparison of FOM_{AMP} for different algorithms as step increases under NMCF. (c)-(d): Radar chart comparison of multiple performance indicators (CMRR ↓, PSRR ↓, Gain ↑, Active Area ↓, FOMS ↑, FOML ↑, Ts ↓) for different methods under NMCF.

Comparison

We aim to compare the proposed framework with existing methods to answer the above two questions. As shown in Table 1, our method consistently achieves the highest FOM_{AMP} values under PVT variations, outperforming all comparative algorithms. The *violation* is defined as the number of performance specifications unmet across PVT corners. Algorithms relying solely on a worst-case-driven approach lack optimization robustness. Their overfocus on extreme scenarios leaves them prone to local optima under dynamic PVT perturbations. Multi-agent frameworks, exemplified by MA-OPT, further exacerbate performance issues in complex environments. Such methods suffer from inconsistent agent coordination and conflicting optimization objectives, which amplify errors and impede convergence. Our method achieves a 170% to 181% improvement in FOM_{AMP} over MA-OPT.

In comparison to Rose and PVTsizing, our method maintains comparable sampling efficiency while achieving superior performance, with average FOM_{AMP} improvements of 34.7% to 44.2% and 13.2% to 20.1%, respectively. Rose's reliance on BO for sampling accelerates RL training but struggles with multi-objective trade-offs and dynamic PVT variations, causing inherent performance losses. PVTsizing, which employs multi-task critics for pruning, lacks precise modeling of PVT-induced performance deviations. In contrast, our method leverages PVT graph representations to explicitly characterize perturbation-induced performance biases, enabling more efficient selective sampling to guide optimization. Additionally, PVT graph-based fine-tuning further reduces violations by dynamically adjusting for perturbation effects. These results confirm that our proposed method achieves a balanced improvement in performance, optimization efficiency, and PVT-robustness.

Case Study

A key advantage of the proposed method is that it uses entropy-regularized rewards to balance exploration and exploitation while aligning with fuzzy priorities and employs a PVT-graph to mitigate reward distribution drift caused by PVT variations. This combination enables faster convergence and stable performance. As shown in Fig.2 (a)-(b), our method achieves higher and more stable FOM_{AMP} in both circuits, outperforming worst-case driven methods like rGNN-RL and MA-OPT. These methods suffer from convergence oscillations induced by PVT variations.

Furthermore, our method addresses reward degradation and diminishing returns in many-objective trade-offs by converting quantitative rewards into qualitative priorities via fuzzy logic. As shown in the radar charts (c)-(d), our method demonstrates superior multi-objective optimization performance across key indicators for both circuits: it ensures consistent policy convergence even with conflicting objectives and under PVT variations, resulting in balanced competitiveness across all metrics. In contrast, other methods often sacrifice performance in specific indicators due to such trade-offs.

Conclusion

This paper proposes a priority-based, graph-enhanced reinforcement learning framework for robust analog circuit optimization under PVT variations. Quantitative rewards are converted into qualitative priority signals via fuzzy logic to alleviate diminishing reward differentiation and unstable policy convergence in many-objective settings. Paired with a physically grounded graph representation that explicitly models correlations among operating conditions, the method efficiently navigates high-dimensional objectives and allocates effort to variation-sensitive regions without sacrificing optimization efficiency. Unlike prior PVT-aware methods that treat scenarios independently or rely on heuristic pruning, this approach captures variation-induced performance deviations more accurately and explores robust design spaces more effectively. Experiments on seven complex circuits show improved solution quality and sample efficiency with fewer constraint violations across PVT corners.

Acknowledgments

The project was supported in part by the Shenzhen Science and Technology Program (Grant No. JCYJ20240813114202004) and in part by the National Natural Science Foundation of China (Grant No. 52405253)

References

Bao, J.; Zhang, J.; Huang, Z.; Bi, Z.; Feng, X.; Zeng, X.; and Lu, Y. 2024. Multiagent Based Reinforcement Learning (MA-RL): An Automated Designer for Complex Analog Circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(12): 4398–4411.

Budak, A. F.; Gandara, M.; Shi, W.; Pan, D. Z.; Sun, N.; and Liu, B. 2022. An Efficient Analog Circuit Sizing Method Based on Machine Learning Assisted Global Optimization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(5): 1209–1221.

Cai, H.; Li, J.; Yu, X.; Zhang, Y.; Luo, T.; Cai, W.; Tang, C.; and Zeng, Y. 2025. Hierarchical multi-task circuit modeling for PVT robustness via KAN-CNN integration. *Expert Systems with Applications*, 274: 126966.

Cao, W.; Gao, J.; Ma, T.; Ma, R.; Benosman, M.; and Zhang, X. 2025. RoSE-Opt: Robust and Efficient Analog Circuit Parameter Optimization With Knowledge-Infused Reinforcement Learning. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 44(2): 627–640.

Chen, Z.; Li, J.; Peng, L.; Li, Y.; Wang, Y.; and Zeng, Y. 2025. Accelerating Comprehensive Specification Optimization of Analog Circuits Using Transient Assertions and Graph Neural Networks. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5. IEEE.

Choi, M.; Choi, Y.; Lee, K.; and Kang, S. 2023. Reinforcement Learning-based Analog Circuit Optimizer using gm/ID for Sizing. In *Proc. DAC*, 1–6.

Choi, Y.; Park, S.; Choi, M.; Lee, K.; and Kang, S. 2024. MA-Opt: Reinforcement Learning-Based Analog Circuit Optimization Using Multi-Actors. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 71(5): 2045–2056.

Ding, Y.; Zhi, H.; Li, J.; Chen, Z.; Yang, K.; and Shan, W. 2023. A Compact and Robust 28nm CMOS Temperature Sensor with Machine Learning Assisted Design for DVFS SoC. In *2023 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA)*, 87–88.

Gao, J.; Cao, W.; and Zhang, X. 2023. RoSE: Robust Analog Circuit Parameter Optimization with Sampling-Efficient Reinforcement Learning. In *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 1–6.

Gielen, G.; Walscharts, H.; and Sansen, W. 1990. Analog circuit design optimization based on symbolic simulation and simulated annealing. *IEEE Journal of Solid-State Circuits*, 25(3): 707–713.

Google; and Foundry, S. T. 2020. SkyWater Open Source PDK for the SKY130 Process Node. https://github.com/google/skywater-pdk.

Gu, T.; Li, W.; Zhao, A.; Bi, Z.; Li, X.; Yang, F.; Yan, C.; Hu, W.; Zhou, D.; Cui, T.; Liu, X.; Zhang, Z.; and Zeng, X. 2024a. BBGP-sDFO: Batch Bayesian and Gaussian Process Enhanced Subspace Derivative Free Optimization for High-Dimensional Analog Circuit Synthesis. *IEEE TCAD*, 43(2): 417–430.

Gu, T.; Wang, J.; Bi, Z.; Yan, C.; Yang, F.; Qin, Y.; Cui, T.; and Zeng, X. 2024b. tSS-BO: Scalable Bayesian Optimization for Analog Circuit Sizing via Truncated Subspace Sampling. In *Proc. DATE*, 1–6.

Hakhamaneshi, K.; Nassar, M.; Phielipp, M.; Abbeel, P.; and Stojanovic, V. 2023. Pretraining Graph Neural Networks for Few-Shot Analog Circuit Modeling and Design. *IEEE TCAD*, 42(7): 2163–2173.

Jain, R.; Xu, S.; Kaushal, R.; Mariscal, C.; Caballero, H.; Salus, T.; Schaef, C.; Deka, A.; Payala, A.; Chen, K.; et al. 2024. 28.6 An 87% Efficient 2V-Input, 200A Voltage Regulator Chiplet Enabling Vertical Power Delivery in Multi-kW Systems-on-Package. In *Proc. ISSCC*, volume 67, 466–468.

Kong, Z.; Tang, X.; Shi, W.; Du, Y.; Lin, Y.; and Wang, Y. 2024. PVTSizing: A TuRBO-RL-Based Batch-Sampling Optimization Framework for PVT-Robust Analog Circuit Synthesis. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, 1–6.

Li, J.; He, S.; Li, A.; Yu, S.; and Li, Y. 2026. Multitask evolution with problem reformulation for global exploration in analog circuit design. *Advanced Engineering Informatics*, 70: 104195.

Li, J.; Zeng, Y.; Zhi, H.; Yang, J.; Shan, W.; Li, Y.; and Li, Y. 2024. Knowledge Transfer Framework for PVT Robustness in Analog Integrated Circuits. *IEEE TCSI*, 71(5): 2017–2030.

Li, J.; Zhi, H.; Lyu, R.; Li, W.; Bi, Z.; Zhu, K.; Zeng, Y.; Shan, W.; Yan, C.; Yang, F.; Li, Y.; and Zeng, X. 2025a. AnalogGym: An Open and Practical Testing Suite for Analog Circuit Synthesis. In *Proceedings of the 43rd IEEE/ACM International Conference on Computer-Aided Design*, ICCAD '24. New York, NY, USA: Association for Computing Machinery. ISBN 9798400710773.

Li, J.; Zhi, H.; Shan, W.; Li, Y.; Zeng, Y.; and Li, Y. 2023a. Multi-Task Evolutionary to PVT Knowledge Transfer for Analog Integrated Circuit Optimization. In *Proc. ICCAD*.

Li, J.; Zhi, H.; Xiao, J.; Zhu, K.; and Li, Y. 2025b. Decoupling Analog Circuit Representation from Technology for Behavior-Centric Optimization. In *2025 62nd ACM/IEEE Design Automation Conference (DAC)*, 1–7.

Li, Y.; Guo, J.; Wang, R.; and Yan, J. 2023b. T2T: From Distribution Learning in Training to Gradient Search in Testing for Combinatorial Optimization. In *Advances in Neural Information Processing Systems*.

Li, Z.; and Carusone, A. C. 2023. Design and Optimization of Low-Dropout Voltage Regulator Using Relational Graph Neural Network and Reinforcement Learning in Open-Source SKY130 Process. In *Proc. ICCAD*, 01–09.

Liu, M.; Turner, W. J.; Kokai, G. F.; Khailany, B.; Pan, D. Z.; and Ren, H. 2021. Parasitic-Aware Analog Circuit Sizing with Graph Neural Networks and Bayesian Optimization. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 1372–1377.

Luo, F.; Lin, X.; Liu, F.; Zhang, Q.; and Wang, Z. 2023. Neural combinatorial optimization with heavy decoder: Toward large scale generalization. In *The 37th Anniversary Conference on Neural Information Processing Systems, NeurIPS 2023*.

Lyu, W.; Xue, P.; Yang, F.; Yan, C.; Hong, Z.; Zeng, X.; and Zhou, D. 2018. An Efficient Bayesian Optimization Approach for Automated Optimization of Analog Circuits. *IEEE TCSI*, 65(6): 1954–1967.

Pan, M.; Lin, G.; Luo, Y.-W.; Zhu, B.; Dai, Z.; Sun, L.; and Yuan, C. 2025. Preference Optimization for Combinatorial Optimization Problems. arXiv:2505.08735.

Shi, W.; Wang, H.; Gu, J.; Liu, M.; Pan, D. Z.; Han, S.; and Sun, N. 2022. RobustAnalog: Fast variation-aware analog circuit design via multi-task RL. In *Proceedings of the 2022 ACM/IEEE Workshop on Machine Learning for CAD*, 35–41.

Sutton, R. S.; Barto, A. G.; et al. 1998. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

Wang, H.; Wang, K.; Yang, J.; Shen, L.; Sun, N.; Lee, H.-S.; and Han, S. 2020. GCN-RL Circuit Designer: Transferable Transistor Sizing with Graph Neural Networks and Reinforcement Learning. In *Proc. DAC*, 1–6.

Wang, H.; Yang, J.; Lee, H.-S.; and Han, S. 2018. Learning to design circuits. *arXiv preprint arXiv:1812.02734*.

Xu, P.; Li, J.; Ho, T.-Y.; Yu, B.; and Zhu, K. 2024. Performance-Driven Analog Layout Automation: Current Status and Future Directions. In *2024 29th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 679–685.

Zhang, J.; Bao, J.; Huang, Z.; Zeng, X.; and Lu, Y. 2023. Automated Design of Complex Analog Circuits with Multiagent based Reinforcement Learning. In *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 1–6.

Zhi, H.; Li, J.; Li, Y.; and Shan, W. 2025a. Analog Circuit Transfer Method Across Technology Nodes via Transistor Behavior. In *Proceedings of the 30th Asia and South Pacific Design Automation Conference*, 197–203.

Zhi, H.; Xu, S.; Li, J.; Zhou, T.; Li, Y.; Shan, W.; and Qu, W. 2025b. Closed-Loop Pole Analysis via Output Impedance in Miller-Compensated Amplifiers. *IEEE Transactions on Circuits and Systems II: Express Briefs*.

Zhong, X.; Li, J.; Bi, Z.; Li, Y.; Yang, F.; Zeng, X.; and Zhu, K. 2025. PZTA: Accelerating Analog Circuit Sizing With A Transferable Circuit Theory-Inspired Pole-Zero Transient Assertion System. In *2025 International Symposium of Electronics Design Automation (ISEDA)*, 169–174. IEEE.