FD-MAGRPO: Functionality-Driven Multi-Agent Group Relative Policy Optimization for Analog-LDO Sizing

Haoning Jiang[*1], **Han Wu**[*1,2], **Zhuoli Ouyang**[1], **Ziheng Wang**[1], **Tinghuan Chen**[3], **Junmin Jiang**[1†]

[1]Southern University of Science and Technology, Shenzhen, China
[2]University of Macau, Macao, China
[3]The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

Abstract

This paper introduces the Functionality-Driven Multi-Agent Group Relative Policy Optimization (FD-MAGRPO) algorithm, which is designed to enhance exploration efficiency in reinforcement learning (RL) for analog integrated circuit sizing. Our proposed method integrates two key innovations: (1) a critic-free multi-agent optimization framework based on Group Relative Policy Optimization (GRPO) that eliminates the critic network and achieves stable and efficient policy updates; and (2) a functionality-driven grouping strategy, that enables agents to coordinate exploration by functional roles instead of circuit blocks, thereby improving credit assignment and cooperation. Experimental results on practical low-dropout regulator (LDO) circuits with 65–179 design parameters show that the proposed method achieves rapid convergence with only 800–3000 simulations, yielding a 4.8×–13.0× speedup over state-of-the-art methods. Mathematical analysis and empirical studies validate that the combination of critic-free optimization and functionality-based grouping leads to higher exploration efficiency and faster convergence. The proposed method enables the discovery of higher circuit performances that are inaccessible to conventional approaches, establishing FD-MAGRPO as a robust and efficient solution for complex analog-LDO sizing tasks.

Introduction

Analog integrated circuits serve as critical interfaces between digital systems and the physical world. Unlike the highly automated workflows in digital circuit design, analog circuit design remains heavily reliant on expert knowledge and manual experience (Barros, Guilherme, and Horta 2010).

Circuit sizing optimization is a critical phase in analog design, where the optimal circuit parameters are determined based on the performance specification. This process is challenging due to the high-dimensional design space, which demands the simultaneous adjustment of numerous parameters to satisfy multiple target specifications and constraints.

Reinforcement learning(RL) serves as a common solution method for analog circuit sizing optimization (Wang et al. 2020; Settaluri et al. 2020; Shi et al. 2022; Settaluri et al. 2022). RL methods provide policy transferability and reusability due to their policy networks, enabling adaptation to new process nodes or target specifications via fine-tuning rather than retraining. Figure 1 highlights the advantages of RL, which not only significantly reduces manual labor costs, but also substantially shortens the overall optimization cycle and accelerates technology migration or target modification.

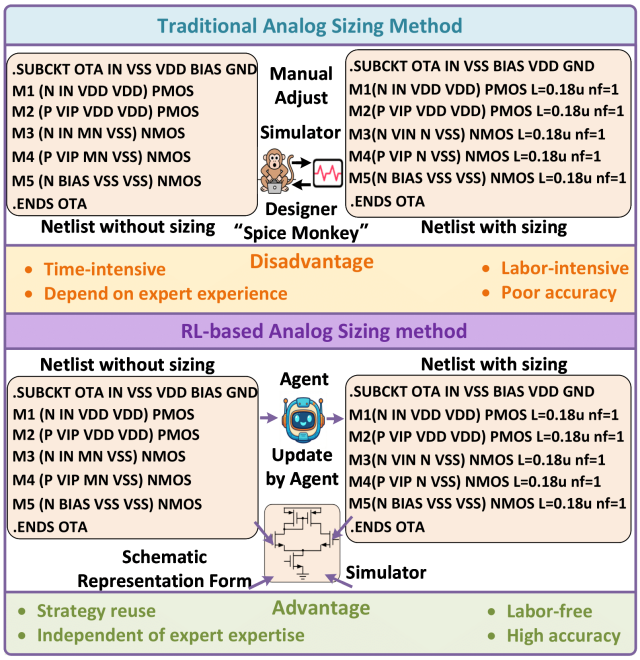

Figure 1: Comparison between traditional analog sizing methods and RL-based approaches. Traditional methods, often referred to as "Spice Monkey," rely on manual adjustments, while RL automates the process, enabling strategy reusability and higher efficiency.

Recent advances in multi-agent reinforcement learning (MARL) show higher sample efficiency than single-agent RL in analog circuit sizing. This is achieved by decomposing complex optimization into coordinated sub-tasks. Homogeneous MARL frameworks using centralized training

[*]These authors contributed equally.
[†]The corresponding author. (Email: jiangjm@sustech.edu.cn)

with decentralized execution (CTDE) and MAPPO algorithms improve convergence rates in analog circuit optimization (Zhang et al. 2023; Bao et al. 2024). Hierarchical MARL architectures further enhance performance by separating system-level objectives from parameter tuning, while integrating Bayesian optimization to reduce simulation costs (Feng et al. 2025). Model-based MARL methods employ probabilistic models for virtual trajectory generation and uncertainty-aware exploration, significantly lowering computational effort while maintaining optimization quality (Ahmadzadeh and Gielen 2024).

Despite these advances, current MARL approaches for analog circuit sizing exhibit two key limitations. First, existing methods rely on actor-critic algorithms ill-suited to the essentially sparse-reward and high-dimensional nature of circuit optimization. Critic networks introduce estimation bias and instability, reducing sample efficiency and slowing convergence in computationally intensive simulation environments (Fujimoto, Hoof, and Meger 2018; Zhou et al. 2021). This is mainly because, in analog circuit sizing, the action space is directly mapped to the entire parameter space, and each episode generates a complete circuit configuration in one step. There are no informative intermediate states and the reward is obtained only after a simulation. Under such conditions, critic networks struggle to accurately estimate value functions, leading to unstable training and inefficient exploration.

Second, prior works typically partition circuits into block-based groups, assigning each agent to specific sub-blocks. However, this grouping strategy misaligns with the functional dependencies inherent in analog circuits. For example, biasing transistors performing identical functional roles may be physically distributed across distinct circuit modules. Block-based grouping thus fails to capture essential functional relationships, which are critical for effective coordination and accurate credit assignment among agents (Yang et al. 2023). This misalignment leads to ambiguous credit assignment, poor cooperation, and ultimately slower convergence during the optimization process in complex analog circuits.

To address these challenges, we propose Functionality-Driven Multi-Agent Group Relative Policy Optimization (FD-MAGRPO), building on Group Relative Policy Optimization (GRPO) (Shao et al. 2024). The main contributions are:

1) GRPO is extended to a multi-agent framework for analog sizing, removing critic networks and ensuring stable policy updates through relative advantage estimation within policy groups. This approach addresses the instability and sample inefficiency caused by critic networks in conventional MARL algorithms, particularly under sparse-reward and high-dimensional circuit optimization scenarios.

2) A novel functionality-driven grouping method is proposed, replacing conventional block-based partitioning. The advantage of this approach is mathematically validated using a vector-level gradient signal-to-noise ratio (GSNR) metric, rigorously quantifying improvements in credit assignment and optimization performance. Further, experimental results validate that the proposed method improves credit assignment and leads to faster convergence.

Preliminaries

This section introduces core concepts for defining the analog circuit sizing problem and the associated RL environment for its optimization.

Formulation of Analog Sizing Optimization

Analog circuit sizing is defined as a constrained optimization problem. Given a parameter vector $\boldsymbol{\phi} \in \mathcal{S}^N$ in an N-dimensional design space (e.g. resistor values, etc) and a performance vector $\boldsymbol{\psi} \in \mathbb{R}^M$ in an M-dimensional specification space (containing metrics like gain, etc), their relationship is defined by $\boldsymbol{\psi} = F(\boldsymbol{\phi})$, where $F(\cdot)$ represents the circuit simulation. The optimization problem is expressed as:

$$\underset{\boldsymbol{\phi}\in\mathcal{S}^N}{\text{maximize}}\ J(\boldsymbol{\psi}) \quad \text{subject to } B(\boldsymbol{\psi}) \leq \mathbf{0} \tag{1}$$

Here, $J(\boldsymbol{\psi})$ is the objective function and $B(\boldsymbol{\psi})$ represents boundary constraints (e.g., power limits, gain thresholds).

Specifications are categorized as follows:

- **Constrained specifications** (K_c): Metrics that must strictly satisfy minimum or maximum requirements to ensure circuit functionality (e.g. phase margin). For these specifications, it is sufficient to meet the constraint; further optimization is unnecessary.
- **Optimal specifications** (K_o): Metrics that can be further optimized to enhance circuit quality once all constraints are satisfied (e.g. bandwidth, efficiency). For these specifications, higher (or lower) values are generally preferred depending on the design objective.

The objective function J directly measures the overall performance. It rewards improved performance only when all specifications are satisfied and penalizes any specification violations.

$$J = \begin{cases} -\alpha \cdot \frac{\sum_{i=1}^{K_c} \lambda_i \min\left(\delta_i^{req},0\right)+\sum_{j=1}^{K_o} \lambda_j \min\left(\delta_j^{req},0\right)}{\sum_{i=1}^{K_c} \lambda_i+\sum_{j=1}^{K_o} \lambda_j} \\ \text{if any target specification is not satisfied} \\ \beta + \gamma \sum_{j=1}^{K_o} \lambda_j \max\left(\delta_j^{opt},0\right) \\ \text{if all target specifications are satisfied} \end{cases} \tag{2}$$

where

$$\delta_k^{req} = \frac{\psi_k - \psi_k^{req}}{\psi_k + \psi_k^{req}}, \qquad \delta_k^{opt} = \frac{\psi_k - \psi_k^{opt}}{\psi_k + \psi_k^{opt}}$$

denote the normalized deviations of specification k from its required value ψ_k^{req} and its optimal target ψ_k^{opt}, respectively. Here, ψ_i^{req} and ψ_j^{req} are the minimum required values for constrained and optimal specifications, respectively, and ψ_j^{opt} is the reference for quantifying performance improvement once all specifications are satisfied. λ_i and λ_j are weights, α is a penalty factor, β is a baseline reward, and γ is a scaling coefficient for the performance bonus.

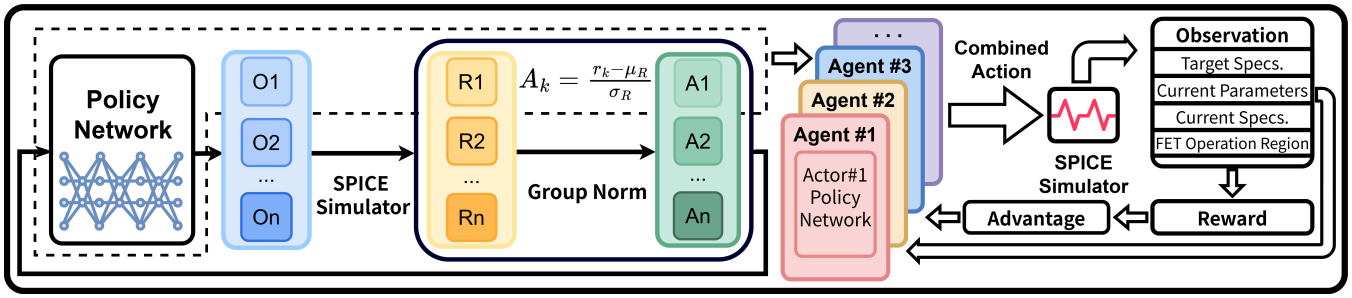

Figure 2: The MAGRPO pipeline assigns agents, each with identical network architectures and independent policies, to distinct functional groups. All agents share a unified observation space that includes target specifications, current parameters, performance metrics, and device operation regions. Agents generate actions corresponding to parameter updates, which are evaluated in parallel by SPICE. Grouped metrics are then normalized to compute advantages.

This formulation treats constraint specifications as a prerequisite for performance optimization. Notably, the objective function provides additional rewards only when all target specifications are met, explicitly incentivizing solutions that exceed minimum requirements.

Custom RL Environment for Analog Sizing

The analog circuit sizing problem is formulated as a RL task through a specialized environment that bridges circuit simulator with RL algorithms. This environment defines how agents interact with circuit simulator by adjusting circuit parameters to achieve optimal performance.

The observation space captures circuit specifications and parameters:

$$\mathcal{O} = [\mathbf{y}_{ideal}, \mathbf{y}_{cur}, \mathbf{P}_{cur}, \mathbf{D}], \tag{3}$$

where $\mathbf{y}_{ideal}$ and $\mathbf{y}_{cur}$ represent target and current specifications normalized by $\psi^{\dagger}$, and $\mathbf{P}_{cur}$ contains circuit parameters:

$$\mathbf{P}_{cur} = [\mathbf{p}_{transistor}, \mathbf{p}_{resistor}, \mathbf{p}_{capacitor}, ...]. \tag{4}$$

For transistors, $\mathbf{p}_{transistor_i} = (w_i, l_i, nf_i)$ represents width, length, and finger count. For resistors and capacitors, $\mathbf{p}_{resistor_i}$ and $\mathbf{p}_{capacitor_i}$ represent the resistance and capacitance values. $\mathbf{D}$ indicates operating regions, ranging from cutoff (0), triode (1), saturation (2), subthreshold (3), to breakdown (4), providing information for working region.

The action space employs continuous normalized values mapped to discrete parameter values through quantization:

$$p_i = p_{min} + \left\lfloor \frac{a_i(p_{max} - p_{min})}{\Delta p} \right\rfloor \Delta p, \tag{5}$$

where $a_i \in [0, 1]$ maps to physical parameter p_i within range $[p_{min}, p_{max}]$ at step size Δp. For an agent controlling k transistors, its action space is:

$$\mathbf{A}^i = [w_1, l_1, nf_1 ... w_k, l_k, nf_k, c_1 ... r_1 ...]. \tag{6}$$

The reward function is formulated to align with the objective function J in Eq. (2). Importance weights λ_i and λ_j for each specification are systematically derived using the Analytic Hierarchy Process (AHP) (Liu, Xu, and Hu 2015), which quantifies the relative significance of each metric.

Methods

Despite recent progress, analog circuit sizing remains fundamentally challenging due to two main factors: (1) the reward signal is extremely sparse and only available after simulating a complete configuration, making critic-based value estimation unreliable and causing unstable training and low sample efficiency; (2) conventional block-based agent grouping fails to capture functional dependencies, resulting in ambiguous credit assignment and poor cooperation. These issues severely limit the convergence speed and optimization quality of current MARL algorithms, especially in high-dimensional and simulation-constrained scenarios.

This section introduces the Functionality-Driven Multi-Agent Group Relative Policy Optimization (FD-MAGRPO) algorithm for analog circuit sizing. The overall architecture is shown in Figure 2.

GRPO and Its Multi-Agent Extension

We strategically adopt single-step RL to achieve rapid convergence in analog circuit sizing optimization. This approach addresses the computational constraints inherent in circuit design workflows. SPICE simulations are computationally intensive, with each evaluation taking seconds to minutes. Maximizing the utilization of each simulation is therefore critical for practical deployment.

Single-step RL enables immediate policy updates that quickly adapt to and explore promising regions of the parameter space. Each simulation result directly informs subsequent parameter selections, creating a tight feedback loop that accelerates convergence. This is particularly valuable for high-dimensional circuit optimization problems where efficient exploration is crucial for discovering optimal solutions within limited simulation budgets.

In such single-step RL scenarios, value estimation by critic networks often proves inaccurate, leading to training instability and inefficient convergence.

To address the limitations of critic-based approaches in analog circuit sizing, we extend GRPO to a multi-agent framework (MAGRPO). This approach eliminates the critic network entirely, replacing it with a more stable and efficient reward normalization mechanism.

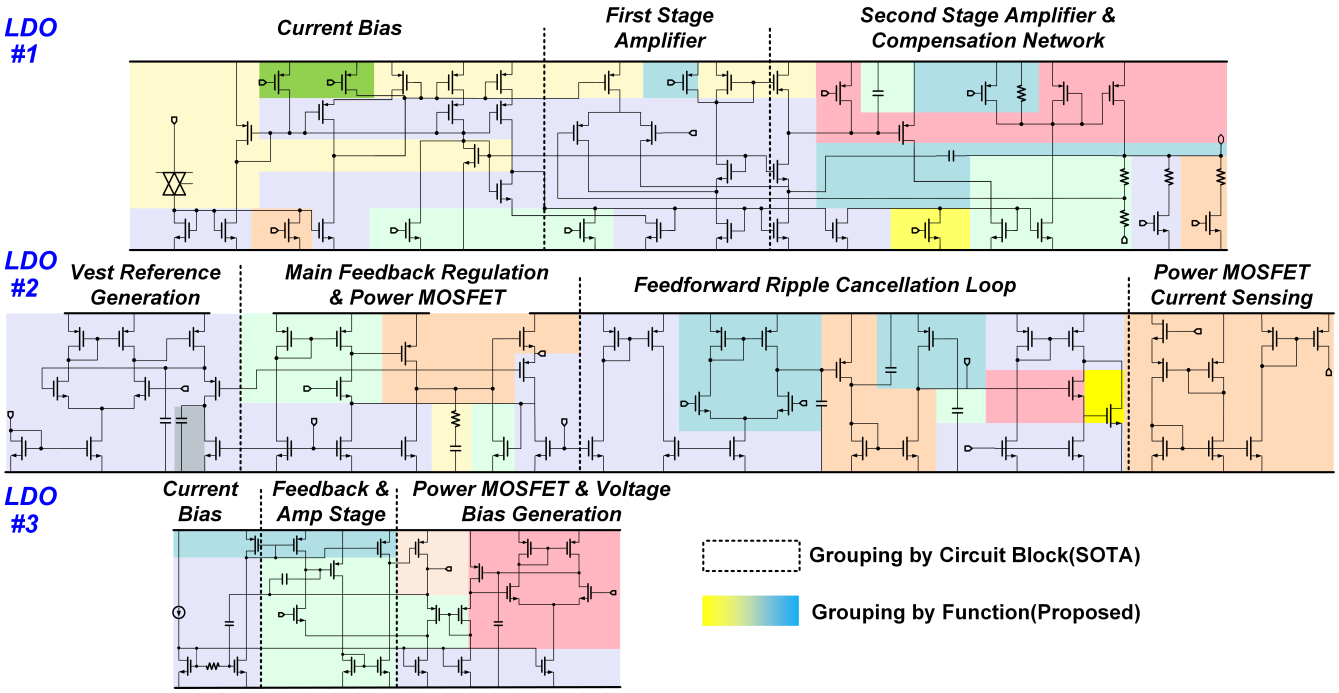

Figure 3: Functionality-based group strategy across different circuit topologies. Different colors represent distinct functionality-based groups, while dashed lines indicate block-based partitions.

The core innovation in GRPO is its direct computation of relative advantages across parallel action batches:

$$\hat{A}_k = \frac{r_k - \mu_R}{\sigma_R} \tag{7}$$

where r_k is the reward of the k-th parallel sample, and μ_R, σ_R are the mean and standard deviation of rewards across the batch. This critic-free normalization efficiently captures relative performance and eliminates estimation bias.

For stable policy updates, GRPO employs a clipped objective function adapted from PPO (Schulman et al. 2017):

$$L^{CLIP}(\theta) = \mathbb{E}_t\big[\frac{1}{K}\sum_{k=1}^{K}\min(\rho_t\hat{A}_k, \text{clip}(\rho_t, 1-\epsilon, 1+\epsilon)\hat{A}_k)\big] \tag{8}$$

where $\rho_t = \frac{\pi_\theta(a^t|s^t)}{\pi_{\text{old}}(a^t|s^t)}$, and ϵ defines the clipping range.

MAGRPO extends this formulation to the multi-agent setting, where circuits are decomposed into distinct functional components, each controlled by an autonomous agent $i \in \mathcal{N}$ with a Gaussian policy. In decentralized mode, all agents share the same reward for each circuit, and advantage normalization is performed over the batch as in Eq. (7). Specifically, the policy of each agent is parameterized as follows:

$$\pi_{\theta_i}(a_i|s_i) = \mathcal{N}(\mu_{\theta_i}(s_i), \sigma_{\theta_i}(s_i)) \tag{9}$$

where $\mu_{\theta_i}(s_i)$ and $\sigma_{\theta_i}(s_i)$ are outputs from the agent's policy network. Final circuit parameters are determined through quantization as described in Eq. (5).

Group Assignment

Analog circuit sizing optimization utilizes a MARL framework, enabling multiple agents to learn concurrently through interactions in a shared environment. This collaborative approach facilitates efficient exploration and experience sharing. However, conventional block-based grouping strategies often fail to capture critical functional relationships among circuit components, limiting their effectiveness in optimizing overall circuit performance. Furthermore, optimizing each circuit block independently does not guarantee optimal global circuit performance, as inter-block dependencies and functional interactions are frequently overlooked.

To address the limitations of block-based grouping, a systematic and well-defined functionality-based grouping strategy is proposed, as illustrated in Figure 3. Agents are assigned according to explicit functional domains that are universally present in analog LDO circuits. Specifically, circuit components are grouped by their functional roles into the following categories: (1) current biasing transistor, (2) voltage biasing transistor, (3) amplifier stage, (4) main feedback loop, (5) auxiliary compensation loop, and (6) power field-effect transistor. Each agent optimizes the parameters of one functional group, regardless of physical location or block affiliation.

This approach enables agents to control units with similar behaviors, resulting in clearer credit assignment and more effective coordination. Subsequent experiments demonstrate that functionality-driven grouping enhances coordination and credit assignment, thereby accelerating convergence speed compared to previous block-based methods.

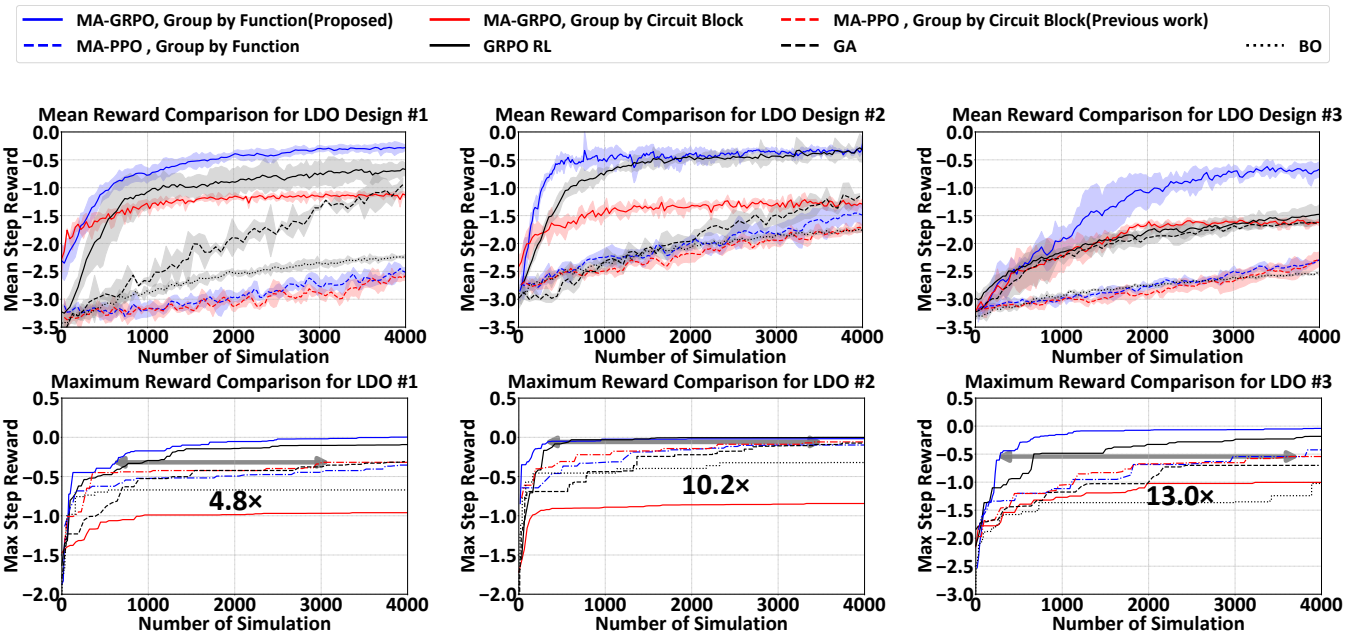

Figure 4: Performance comparison of algorithms across three LDO circuit topologies. The top row shows the average reward progression under simulation counts for each LDO circuit, while the bottom row presents the maximum reward progression. Each column corresponds to a specific LDO circuit.

Evaluations

Experiment Setup

The proposed FD-MAGRPO method was validated by three practical analog LDO circuits, as shown in Figure 3. Performance comparisons were conducted against the state-of-the-art MARL algorithm used in circuit sizing (Bao et al. 2024).

All experiments were deployed on a server equipped with Intel Xeon Gold 6240 processor (2.6 GHz) with 128 GB RAM. Circuit simulations were executed within the Cadence Spectre Simulator using commercial 65 nm and 55 nm CMOS process design kits (PDK).

Research Questions and Experimental Design

This evaluation investigates two central research questions:

1. **Impact of Grouping Strategy:** How does functionality-based grouping compare to block-based grouping in circuit sizing optimization?
2. **Critic-Free vs. Actor-Critic:** How does the critic-free MAGRPO algorithm perform relative to standard actor-critic methods such as MAPPO?

Five algorithm configurations were implemented:

- **MAGRPO + Function Grouping** (proposed): Integrates MAGRPO with functionality-based grouping.
- **MAGRPO + Block Grouping**: Applies MAGRPO with block-based grouping to isolate the effect of grouping strategy.
- **GRPO**: Utilizes single-agent GRPO to assess the benefits of multi-agent design.
- **MAPPO + Function Grouping**: Employs the MAPPO algorithm (Bao et al. 2024) with functionality-based grouping to highlight the critic-free approach.
- **MAPPO + Block Grouping**: The baseline method, im-

Specification	Reported Specs.	Optimized Specs.	Improv.
IQ(μA)	8.00	7.91	**1%** ↑
Line Reg.[1](mV/V)	4.50	1.80	**60%** ↑
Line Reg.[2](mV/V)	7.80	1.96	**75%** ↑
Load Reg.(V/A)	0.12	0.0021	**98%** ↑
Overshoot[3](V)	0.11	0.092	**16%** ↑
Undershoot[3](V)	0.077	0.065	**16%** ↑
Overshoot[4](V)	0.072	0.072	**0%** ↑
Undershoot[4](V)	0.062	0.059	**5%** ↑
PSR@100Hz[2](dB)	51.00	54.85	**8%** ↑
PSR@1kHz[2](dB)	45.00	53.66	**19%** ↑
PSR@10kHz[2](dB)	25.00	40.45	**62%** ↑
PSR@100kHz[2](dB)	3.00	19.27	**542%** ↑
PSR@1MHz[2](dB)	0	7.32	**∞%** ↑
PSR@100Hz[1](dB)	50.00	54.57	**9%** ↑
PSR@1kHz[1](dB)	45.00	53.45	**19%** ↑
PSR@10kHz[1](dB)	25.00	40.48	**62%** ↑
PSR@100kHz[1](dB)	5.00	19.27	**285%** ↑
PSR@1MHz[1](dB)	0	7.51	**∞%** ↑
FoM(fFV/A)	0.588	0.581	**1%** ↑

[1]ILOAD=1mA, [2]ILOAD=100mA, [3]VDD=0.75V, [4]VDD=1.2V.
Bold: Better than reported specifications.

Table 1: Summary of Optimization Results with LDO #3

plemented as in (Bao et al. 2024).

All algorithm configurations were evaluated using identical hyperparameter settings for each circuit topology. The same hyperparameter configuration was applied across all experiments within each algorithm category to ensure fair comparison. Key parameters, such as the number of learning episodes, parallel environments, and neural network architectures, were kept consistent throughout the evaluation.

Functionality-Based Grouping Enhances Credit Assignment

Previous work has applied the gradient signal-to-noise ratio (GSNR) at the parameter level to analyze generalization performance in single-model settings (Liu, Wang, and Zhou 2020; Michalkiewicz et al. 2023). These approaches, however, do not address multi-agent credit assignment problems, where the clarity of each agent's learning signal is critical for convergence speed.

To assess the impact of functionality-based grouping on credit assignment, we compare two partitions of the circuit parameter set $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^{n}$:

$$\mathcal{G}_M = \{G_1^M, \ldots, G_K^M\}, \quad \mathcal{G}_F = \{G_1^F, \ldots, G_K^F\}, \quad (10)$$

where G_k^M groups the parameters by physical block and G_k^F by functional role. Prior work in MARL suggests that coherent grouping facilitates clearer credit assignment and faster convergence (Foerster et al. 2016; Sukhbaatar and Fergus 2016; Foerster et al. 2018).

We quantify the clarity of the learning signal of agent i using the per-agent GSNR:

$$\mathrm{GSNR}_i = \frac{\left\|\mathbb{E}_t[\nabla_{\theta^i} J^{(t)}]\right\|_2}{\sqrt{\mathbb{E}_t\left[\|\nabla_{\theta^i} J^{(t)} - \mathbb{E}_t[\nabla_{\theta^i} J^{(t)}]\|_2^2\right] + \epsilon}}, \quad (11)$$

where $\nabla_{\theta^i} J^{(t)}$ denotes the stochastic policy gradient of agent i. We report both arithmetic and harmonic means across all R agents:

$$\overline{\mathrm{GSNR}} = \frac{1}{R}\sum_{i=1}^{R} \mathrm{GSNR}_i \,, \overline{\mathrm{GSNR}}_h = \left(\frac{1}{R}\sum_{i=1}^{R} \frac{1}{\mathrm{GSNR}_i + \epsilon}\right)^{-1} \quad (12)$$

The arithmetic mean $\overline{\mathrm{GSNR}}$ reflects the typical update quality, whereas the harmonic mean $\overline{\mathrm{GSNR}}_h$ penalizes agents with low GSNR. Table 2 reports these metrics for both grouping methods across three LDO topologies. Functionality-based grouping yields higher GSNR during early exploration, which reduces gradient variance and improves credit assignment, thereby accelerating convergence.

Results and Analysis

As shown in Figure 4 and Table 1, the results highlight two main findings:

1. **Functionality-Based Grouping Improves Credit Assignment:** Functionality-based grouping consistently surpasses block-based grouping across all algorithms. This advantage is attributed to enhanced credit assignment and reduced gradient variance, as measured by the gradient signal-to-noise ratio (GSNR) in MARL algorithms.
2. **Critic-Free MAGRPO Accelerates Convergence:** The critic-free MAGRPO algorithm achieves faster convergence and higher final rewards compared to actor-critic MAPPO. By removing critic-induced bias and variance, MAGRPO produces more accurate gradient estimates and improves sample efficiency.

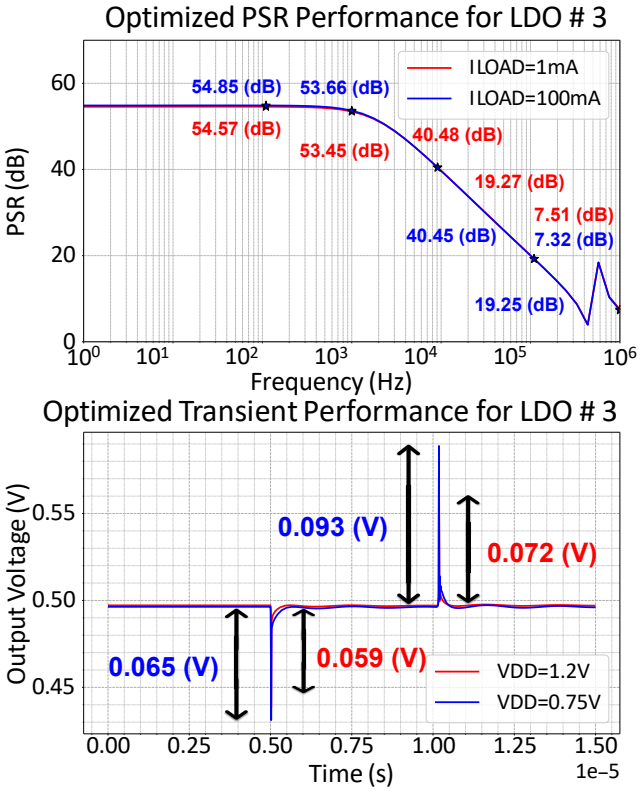

Figure 5: The PSR and transient response simulation results of LDO #3. (a) PSR simulation (Up); (b) Transient simulation (Down).

For fair comparison, the Figure-of-Merit (FoM) (Mao, Lu, and Martins 2022) is adopted to quantitatively evaluate optimization results:

$$\mathbf{FoM} = \frac{I_Q C_L \Delta V_{\mathbf{OUT}}}{\Delta I_{\mathbf{LOAD}}^2}, \quad (13)$$

where I_Q is the quiescent current, C_L is the load capacitance, $\Delta V_{\mathbf{OUT}}$ denotes output voltage deviation during load transients, and $\Delta I_{\mathbf{LOAD}}$ is the load transient current. Table 1 provides a detailed comparison of reported (Guo and Leung 2010) and optimized specifications for one representative LDO #3. The FoM is also significantly reduced, confirming the effectiveness and robustness of FD-MAGRPO in practical analog circuit sizing optimization. The definitions and evaluation criteria for all performance metrics in Table 1 are provided in the Appendix.

Step	LDO #1				LDO #2				LDO #3			
	Functionality-based		Block-based		Functionality-based		Block-based		Functionality-based		Block-based	
	$\overline{\text{GSNR}}$	$\overline{\text{GSNR}}_h$	$\overline{\text{GSNR}}$	$\overline{\text{GSNR}}_h$	$\overline{\text{GSNR}}$	$\overline{\text{GSNR}}_h$	$\overline{\text{GSNR}}$	$\overline{\text{GSNR}}_h$	$\overline{\text{GSNR}}$	$\overline{\text{GSNR}}_h$	$\overline{\text{GSNR}}$	$\overline{\text{GSNR}}_h$
1	0.222 (**+11.6%**)	0.211 (**+7.1%**)	0.199	0.197	0.227 (**+2.7%**)	0.219 (**+5.3%**)	0.221	0.208	0.232 (**+17.8%**)	0.224 (**+18.5%**)	0.197	0.189
5	0.198 (**+15.8%**)	0.186 (**+10.7%**)	0.171	0.168	0.223 (**+10.4%**)	0.216 (**+6.9%**)	0.202	0.202	0.198 (**+33.8%**)	0.196 (**+35.2%**)	0.148	0.145
10	0.248 (**+15.9%**)	0.247 (**+18.8%**)	0.214	0.208	0.225 (**+13.1%**)	0.220 (**+11.7%**)	0.199	0.197	0.224 (**+1.8%**)	0.219 (**+2.3%**)	0.220	0.214
20	0.293 (**+35.0%**)	0.272 (**+32.7%**)	0.217	0.205	0.245 (**+26.3%**)	0.239 (**+25.8%**)	0.194	0.190	0.244 (**+16.7%**)	0.238 (**+17.2%**)	0.209	0.203
30	0.278 (**+30.5%**)	0.273 (**+41.5%**)	0.213	0.193	0.213 (**+8.7%**)	0.207 (**+7.3%**)	0.196	0.193	0.223 (**+42.0%**)	0.217 (**+40.0%**)	0.157	0.155
40	0.296 (**+44.4%**)	0.283 (**+52.2%**)	0.205	0.186	0.215 (**+5.4%**)	0.209 (**+11.8%**)	0.204	0.187	0.226 (**+20.9%**)	0.219 (**+19.0%**)	0.187	0.184

Note: $\overline{\text{GSNR}}_h$ denotes the harmonic mean GSNR. **Bold percentages** indicate the improvement of the proposed grouping method over the current grouping method in terms of GSNR optimization.

Table 2: Gradient SNR (steps 1, 5, 10, 20, 30, 40) for Functionality-based vs. Block-based Grouping over Three LDO Topologies. Functionality-based grouping achieves consistently higher $\overline{\text{GSNR}}$ and $\overline{\text{GSNR}}_h$.

Work	(Ahmadzadeh and Gielen 2024)		(Bao et al. 2024)			**This work**		
	#1	#2	#1	#2	#3	**#1**	**#2**	**#3**(Guo and Leung 2010)
Number of Parameters	20	48	21	42	28	**179**	**149**	**65**
Parameter Space Size	-	-	-	-	-	1.1×10^{204}	5.7×10^{152}	2.5×10^{90}
Number of Performance Specifications	8	8	5	5	6	**14**	**23**	**18**
Simulation Count	1.5k	1.4k	15k	8k	10k	**2k**	**0.8k**	**3k**

Table 3: Performance comparison of analog circuit sizing methods with state-of-the-art works.

The optimized circuit's PSR and load transient responses are shown in Figure 5. These results demonstrate that the optimized sizing ensures correct circuit operation and full compliance with all specified performance requirements.

These comparisons collectively demonstrate the advantages of the proposed FD-MAGRPO method over existing approaches. The integration of functionality-driven grouping and a critic-free architecture addresses key limitations of prior methods, enabling more efficient and effective analog circuit sizing optimization workflows.

Comparison with State-of-the-Art Methods

Table 3 highlights the efficiency of FD-MAGRPO in complex circuit optimization tasks. The validation circuits contain significantly more parameters and larger parameter spaces than those addressed by previous state-of-the-art MARL methods (Ahmadzadeh and Gielen 2024; Bao et al. 2024). Despite this increased complexity, FD-MAGRPO converges with far fewer simulations, demonstrating superior sample efficiency and scalability for industrial-scale analog sizing.

Conclusion and Future Work

In this paper, we present Functionality-Driven Multi-Agent Group Relative Policy Optimization (FD-MAGRPO), an advanced method for analog circuit sizing optimization that addresses key limitations of existing reinforcement learning techniques. The proposed approach introduces two principal innovations: a critic-free architecture for direct advantage estimation, and a functionality-driven grouping method. Both the mathematical analysis and experimental results demonstrate the superiority of FD-MAGRPO in terms of optimization efficiency and solution quality.

Experimental results demonstrate that our method delivers superior performance with substantially reduced computational cost. Only 800–3000 simulations are required, yielding a 4.8×–13.0× speedup over previous state-of-the-art methods. This efficiency becomes increasingly important as circuit complexity grows.

Future research will focus on developing adaptive grouping mechanisms to dynamically identify optimal functional relationships during training, extending MAGRPO to more types of circuits and other complex engineering domains beyond analog sizing.

Acknowledgments

This work was supported by Guangdong Basic and Applied Basic Research Foundation (2025A1515010698), by Shenzhen Fundamental Research Program (JCYJ20240813100601003), by State Key Laboratory of Radio Frequency Heterogeneous Integration (Open Scientific Research Program No. KF2025016), and by High-level Special Funds of SUSTech (G03034K007).

References

Ahmadzadeh, M.; and Gielen, G. G. E. 2024. Using Probabilistic Model Rollouts to Boost the Sample Efficiency of Reinforcement Learning for Automated Analog Circuit Sizing. In *ACM/IEEE Design Automation Conference*, 1–6.

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.

Barros, M.; Guilherme, J.; and Horta, N. 2010. Analog Circuits Optimization Based on Evolutionary Computation Techniques. *Integration, the VLSI Journal*, 43(1): 136–155.

Feng, X.; Xu, Y.; Huang, Z.; Xu, W.; Bi, Z.; Yang, F.; Zeng, X.; and Lu, Y. 2025. Hierarchical Integration of Reinforcement Learning and Optimization Algorithms for Time-Efficient Design Automation of Complex Analog Circuits. *ACM Transactions on Design Automation of Electronic Systems*, 30(3): 42.

Foerster, J.; Nardelli, N.; Farquhar, G.; Torr, P. H. S.; Kohli, P.; and Whiteson, S. 2018. Counterfactual Multi-Agent Policy Gradients. In *AAAI Conference on Artificial Intelligence*.

Foerster, J. N.; Assael, Y. M.; de Freitas, N.; and Whiteson, S. 2016. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:1605.06676*.

Fujimoto, S.; Hoof, H.; and Meger, D. 2018. Addressing Function Approximation Error in Actor-Critic Methods. *arXiv preprint arXiv:1802.09477*.

Guo, J.; and Leung, K. N. 2010. A 6-μW Chip-Area-Efficient Output-Capacitorless LDO in 90-nm CMOS Technology. *IEEE Journal of Solid-State Circuits*, 45(9): 1896–1905.

Liu, C.; Xu, X.; and Hu, D. 2015. Multiobjective Reinforcement Learning: A Comprehensive Overview. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(3): 385–398.

Liu, X.; Wang, Y.; and Zhou, Z. 2020. Understanding Why Neural Networks Generalize Well Through GSNR of Parameters. In *International Conference on Learning Representations*.

Mao, X.; Lu, Y.; and Martins, R. P. 2022. A 1.2-A Calibration-Free Hybrid LDO With In-Loop Quantization and Auxiliary Constant Current Control Achieving High Accuracy and Fast DVS. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(11): 4443–4452.

Michalkiewicz, M.; et al. 2023. Domain Generalization Guided by Gradient Signal to Noise Ratio of Parameters. In *IEEE/CVF International Conference on Computer Vision*, 6177–6188.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*.

Settaluri, K.; Haj-Ali, A.; Huang, Q.; Hakhamaneshi, K.; and Nikolic, B. 2020. AutoCkt: Deep Reinforcement Learning of Analog Circuit Designs. In *Design, Automation and Test in Europe Conference and Exhibition*, 490–495.

Settaluri, K.; Liu, Z.; Khurana, R.; Mirhaj, A.; Jain, R.; and Nikolic, B. 2022. Automated Design of Analog Circuits Using Reinforcement Learning. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(9): 2794–2807.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y. K.; Wu, Y.; et al. 2024. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. *arXiv preprint arXiv:2402.03300*.

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 *ACM/IEEE Workshop on Machine Learning for CAD (MLCAD)*, 35–41.

Sukhbaatar, S.; and Fergus, R. 2016. Learning Multiagent Communication with Backpropagation. *arXiv preprint arXiv:1605.07736*.

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 *ACM/IEEE Design Automation Conference*, 1–6.

Yang, Z.; Jin, H.; Ding, R.; You, H.; Fan, G.; Wang, X.; and Zhou, C. 2023. DeCOM: Decomposed Policy for Constrained Cooperative Multi-Agent Reinforcement Learning. In *AAAI Conference on Artificial Intelligence*, 1219–1228.

Zhang, J.; Bao, J.; Huang, Z.; Zeng, X.; and Lu, Y. 2023. Automated Design of Complex Analog Circuits with Multi-agent Based Reinforcement Learning. In *ACM/IEEE Design Automation Conference*, 1–6.

Zhou, B.; Li, K.; Zeng, H.; Wang, F.; and Tian, H. 2021. ADER: Adapting Between Exploration and Robustness for Actor-Critic Methods. *arXiv preprint arXiv:2109.03443*.