ACEMARL: Adaptive Clustering Enhanced Multi-Agent Reinforcement Learning for Analog Circuit Sizing

Han Wu*[1,2], Haoning Jiang*[1], Zhuoli Ouyang[1], Ziheng Wang[1], Qi Shen[1], Bo Yuan[1], Yan Lu[3], and Junmin Jiang[1†]
[1]Southern University of Science and Technology, Shenzhen, China
[2]University of Macau, Macau, China [3]Tsinghua University, Beijing, China
*Equally Credited Authors (ECAs) †Email: jiangjm@sustech.edu.cn

***Abstract*—Analog circuit sizing remains a critical bottleneck in integrated circuit design, requiring extensive manual effort and computational resources. While multi-agent reinforcement learning (MARL) accelerates optimization through parallel agent training, existing approaches rely on manual circuit block clustering that fails to capture functional relationships between parameters. This paper presents ACEMARL, an adaptive clustering framework that automatically discovers functionally similar parameter clusters. ACEMARL integrates Bi-population Covariance Matrix Adaptation Evolution Strategy (BIPOP-CMA-ES), a high-performance evolutionary algorithm, for multi-modal exploration with data-driven clustering, aiming for automatic agent assignment. Experimental validation on amplifier and low-dropout regulators with up to 179 parameters demonstrated 3.3-5.0× faster convergence and 5.7-38.5% Figure-of-Merit (FoM) improvement compared to state-of-the-art (SOTA) block-based methods. The framework reduced confidence interval width by 31.6-60.7% along mean reward trajectories, enabling fully automated analog circuit sizing with improved stability and performance.**

***Index Terms*—analog circuit sizing, multi-agent reinforcement learning, automated clustering, evolutionary algorithms, BIPOP-CMA-ES**

I. Introduction

Analog circuit sizing determines optimal component parameters to meet performance specifications, requiring extensive expertise and computational resources. Engineers must repeatedly optimize circuits when specifications or processes change. The increasing complexity of modern circuits demands automated design methodologies.

Reinforcement learning (RL) enables analog circuit automation through generalization: trained policies adapt to different specifications and processes without retraining [1]. Single-agent RL suffers from low sample efficiency in high-dimensional parameter spaces [2]. Multi-agent reinforcement learning (MARL) decomposes optimization across specialized agents for faster convergence. Existing MARL implementations rely on manual block-based clustering, assigning parameters by circuit modules rather than functional relationships [3]. This structural clustering misaligns with functional similarity, limiting MARL performance.

Automated clustering based on functional relationships is essential for MARL in circuit sizing. This paper presents ACEMARL, an Adaptive Clustering Enhanced Multi-Agent Reinforcement Learning framework that discovers parameter clusters through data-driven analysis. ACEMARL employs BIPOP-CMA-ES for high-value sampling, correlation analysis for parameter relationships, and automatic agent assignment. This approach eliminates manual intervention while improving optimization performance.

The main contributions are:

- An automated clustering framework discovering functional parameter relationships without expert knowledge.
- Integration of BIPOP-CMA-ES with correlation analysis for high-value sampling and automatic clustering.
- Validation on circuits with 85-179 parameters achieving 3.3-5.0× faster convergence and 31.6-60.7% decrease in confidence interval width than state-of-the-art methods.

II. Related Works

A. Previous Researches

Various approaches have been proposed for automating analog circuit sizing. Optimization-based methods include genetic algorithms [4] [5] and Bayesian optimization [6] [7] [8]. These methods require complete re-optimization when specifications change. This limitation motivates learning-based approaches with generalization capabilities.

Learning-based methods enable generalization. Supervised learning [9] maps specifications to parameters but requires extensive labeled datasets. RL [10] [11] learns optimization policies through circuit simulation feedback, enabling adaptation to new specifications without retraining. RL demonstrates promise for complex analog designs [1], though sample efficiency remains a critical challenge [2]. Multi-agent approaches address this efficiency challenge through parallel optimization.

MARL improves efficiency by decomposing the optimization across specialized agents. Zhang et al. [3] applied MARL to analog sizing using manual block-based clustering, assigning parameters by circuit modules. However, this clustering misaligns with functional relationships. Hybrid approaches [12]

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). *(Han Wu and Haoning Jiang contributed equally to this work.) (Corresponding author: Junmin Jiang)*

978-3-9826741-1-7/DATE26/© 2026 EDAA

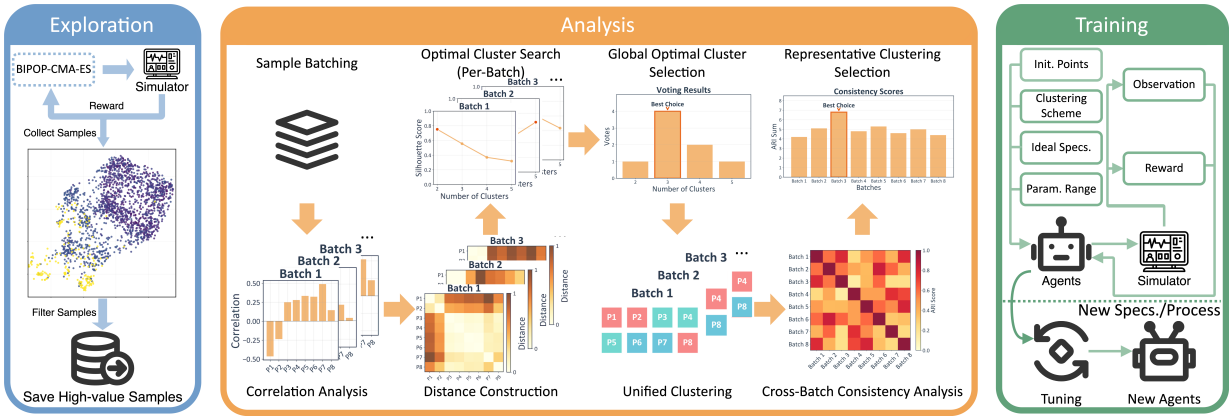

Fig. 1. Overview of the ACEMARL framework. Three-stage pipeline: (1) Exploration phase using BIPOP-CMA-ES to generate high-value samples through adaptive population strategies; (2) Analysis phase performing correlation-based parameter clustering with multi-batch validation to identify functionally similar parameter clusters; (3) Training phase implementing MARL with automatically assigned agents based on clustering results.

combine evolutionary algorithms with RL for initialization but lack structural insight extraction.

B. Research Gaps

Automated parameter clustering for MARL in analog circuit sizing remains unexplored. Existing methods rely on circuit block clustering with designer knowledge. Components performing similar functions should be clustered together [13]. However, functional components (e.g., signal amplification, biasing, pole-zero adjustment) may span different blocks while same-block components may serve distinct functions. This mismatch limits scalability and optimization performance. These critical limitations remain unaddressed in the previous researches, preventing MARL from practical application in analog circuit sizing. ACEMARL addresses this gap through data-driven clustering that discovers functional relationships from optimization samples.

III. Problem Formulation

The analog circuit sizing problem is formulated as constrained multi-objective optimization. Given design parameters $\mathbf{x} \in \mathcal{D}^N$ with feasible space $\mathcal{D}$ and dimension N, performance metrics $\mathbf{y} \in \mathbb{R}^M$ are obtained through SPICE simulation: $\mathbf{y} = f(\mathbf{x})$.

The optimization objective is:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{D}^N} \mathcal{R}(\mathbf{y}) \quad \text{s.t.} \quad \mathcal{C}(\mathbf{y}) \leq 0, \tag{1}$$

where $\mathcal{R}(\mathbf{y})$ denotes the reward function, and $\mathcal{C}(\mathbf{y})$ represents design constraints.

The reward function employs two-stage evaluation:

$$\mathcal{R} = \begin{cases} \alpha \cdot \dfrac{\sum_{i=1}^{N_s} \omega_i \cdot \min\left(\frac{y_i - y_i^{\text{target}}}{y_i + y_i^{\text{target}}}, 0\right) + \sum_{j=1}^{N_o} \omega_j \cdot \min\left(\frac{y_j - y_j^{\text{target}}}{y_j + y_j^{\text{target}}}, 0\right)}{\sum_{i=1}^{N_s} \omega_i + \sum_{j=1}^{N_o} \omega_j} \\ \quad \text{if baseline requirements unmet} \\ \beta + \gamma \cdot \sum_{j=1}^{N_o} \omega_j \cdot \max\left(\frac{y_j - y_j^{\text{target}}}{y_j + y_j^{\text{target}}}, 0\right), \\ \quad \text{if all baseline requirements met} \end{cases} \tag{2}$$

where N_s denotes satisfactory specifications requiring minimum thresholds, N_o represents optimization specifications for continuous improvement, and weights ω_i, ω_j reflect relative importance. The multi-modal optimization landscape motivates the proposed three-stage framework.

IV. Proposed ACEMARL Framework

A. Framework Overview

ACEMARL employs a three-stage pipeline integrating evolutionary exploration, correlation-based clustering, and MARL training, as illustrated in Fig. 1. BIPOP-CMA-ES generates high-value samples for both correlation analysis and RL initialization. Correlation analysis discovers functional parameter clusters from high-value samples. MARL agents specialize in their automatically assigned clusters, achieving faster convergence than block-based clustering.

B. High-value Sample Generation with BIPOP-CMA-ES

Analog circuit optimization landscapes exhibit multi-modal characteristics, as demonstrated by Uniform Manifold Approximation and Projection (UMAP) [14] visualization of historical optimization data from a 149-parameter circuit in Fig. 2. High-value samples distribute across disjoint regions rather than forming single clusters, necessitating sophisticated exploration strategies beyond local search methods.

BIPOP-CMA-ES [15] addresses this challenge through adaptive restart strategies detailed in Algorithm 1. The algorithm alternates between large populations with increasing size ($\lambda = \lambda_0 \cdot 2^{n_L}$) for exploitation and small populations ($\lambda = \max(4, \lambda_0/3)$) with randomized step sizes ($\sigma = 2 \cdot \text{rand}()^2$) for exploration. This dual-population approach ensures comprehensive coverage of the parameter space.

Validation on the Rastrigin function demonstrated effectiveness: while standard CMA-ES converges to local optima (minimum value achieved: 3.98), BIPOP-CMA-ES locates near the global optimum (minimum value achieved: 0.008), as shown in Fig. 3. For circuit optimization, the exploration phase generates 2,00-4,000 samples, with the top 10-50% selected as high-value samples $\mathcal{S}_{\text{high}}$.

Algorithm 1: BIPOP-CMA-ES for High-Value Sample Generation

Data: Action dimension N, initial population size λ_0, initial step size σ_0, max restarts $r_{\max}$
Result: High-value samples $\mathcal{S}$ and best solution $\mathbf{x}^*$
$\mathcal{S} \leftarrow \emptyset$; $r \leftarrow 0$; $\mathcal{R}^* \leftarrow -\infty$; $n_L \leftarrow 0$; $n_S \leftarrow 0$;
Initialize CMA-ES with $\mathbf{m} \leftarrow [0.5]^N$, $\sigma \leftarrow \sigma_0$, $\lambda \leftarrow \lambda_0$;
while $r < r_{\max}$ *and not converged* **do**
  $\mathbf{X} \leftarrow$ Sample λ solutions from CMA-ES;
  for $i = 1$ **to** λ **do**
    $\mathcal{R}_i \leftarrow$ Evaluate($\mathbf{x}_i$); $\mathcal{S} \leftarrow \mathcal{S} \cup \{(\mathbf{x}_i, \mathcal{R}_i)\}$;
  end
  Update $\mathbf{x}^*$, $\mathcal{R}^*$ and CMA-ES distribution with $(\mathbf{X}, -\mathcal{R})$;
  if *restart condition met* **then**
    $r \leftarrow r + 1$;
    if $n_L \leq n_S$ **then**
      $n_L \leftarrow n_L + 1$; $\lambda \leftarrow \lambda_0 \cdot 2^{n_L}$; $\sigma \leftarrow \sigma_0$; // Large pop.
    else
      $n_S \leftarrow n_S + 1$; $\lambda \leftarrow \max(4, \lambda_0/3)$;
      $\sigma \leftarrow 2 \cdot \text{rand}()^2$; // Small pop.
    end
    Reinitialize CMA-ES with new λ, σ;
  end
end
return *FilterTopSamples*($\mathcal{S}$, *threshold*), $\mathbf{x}^*$

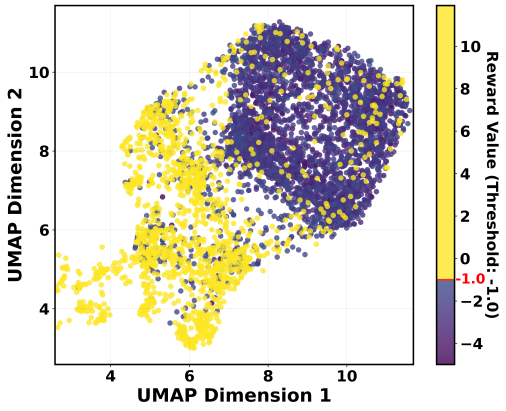

Fig. 2. UMAP visualization of 149-parameter circuit historical optimization data. High-value samples ($\mathcal{R} > -1.0$, shown in yellow) distribute across disjoint regions, demonstrating the multi-modal optimization landscape.

C. Automated Parameter Clustering

The core innovation of ACEMARL lies in automated discovery of parameter relationships through correlation analysis, detailed in Algorithm 2. In analog circuits, functionally related parameters exhibit statistical correlations due to physical coupling. During optimization, these parameters demonstrate coordinated changes. Specifically, when one parameter increases, functionally related parameters adjust correspondingly to maintain performance. The optimization trajectories captured in high-value samples encode these functional relationships through correlation patterns. Parameters serving similar functions (e.g., biasing transistors across different blocks) show similar correlation profiles with rewards, while functionally distinct parameters exhibit different patterns.

Algorithm 2: Correlation-Based Parameter Clustering

Data: High-value samples $\mathcal{S}_{\text{high}}$, batch count M, batch size N
Result: Parameter clustering assignment $\mathbf{C}^*$
Randomly partition $\mathcal{S}_{\text{high}}$ into M batches of size N;
for $m = 1$ **to** M **do**
  $\mathbf{r}^{(m)} \leftarrow \text{PearsonCorr}(\mathbf{X}_m, \mathcal{R}_m)$;
  $\mathbf{D}^{(m)}[i,j] \leftarrow |\mathbf{r}_i^{(m)} - \mathbf{r}_j^{(m)}| \ \forall i,j$;
  for $k = 2$ **to** $K_{\max}$ **do**
    $\mathbf{C}_k^{(m)} \leftarrow \text{HierarchicalCluster}(\mathbf{D}^{(m)}, k)$;
    $S_k^{(m)} \leftarrow \text{SilhouetteScore}(\mathbf{D}^{(m)}, \mathbf{C}_k^{(m)})$;
  end
  $k_{\text{best}}^{(m)} \leftarrow \arg\max_k S_k^{(m)}$;
end
$K^* \leftarrow \text{Mode}(\{k_{\text{best}}^{(1)}, \ldots, k_{\text{best}}^{(M)}\})$;
for $m = 1$ **to** M **do**
  $\mathbf{C}^{(m)} \leftarrow \text{HierarchicalCluster}(\mathbf{D}^{(m)}, K^*)$;
end
$\mathbf{A}[i,j] \leftarrow \text{ARI}(\mathbf{C}^{(i)}, \mathbf{C}^{(j)}) \ \forall i,j$;
$m^* \leftarrow \arg\max_m \sum_{j \neq m} \mathbf{A}[m,j]$;
return $\mathbf{C}^{(m^*)}$

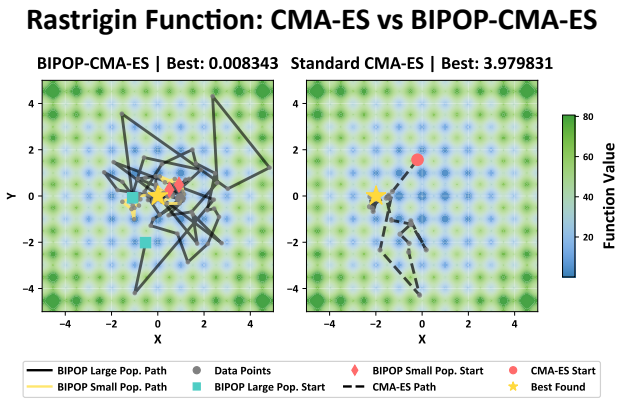

Fig. 3. Comparison of optimization trajectories on Rastrigin function: Standard CMA-ES converges to local optimum; BIPOP-CMA-ES explores multiple regions through restarts (black: large population, yellow: small population), successfully finding the global optimum.

The clustering pipeline employs multi-batch validation for statistical robustness. Samples $\mathcal{S}_{\text{high}}$ are randomly partitioned into M batches (typically 4-8) of size N (32-256 samples). For each batch, Pearson correlations are computed between parameters and rewards. Transistors with multiple design variables (width w, length l, finger count nf) are represented by the composite metric $w \cdot nf/l$ to capture overall performance impact. The correlation-based distance metric $d_{ij} = |\rho_i - \rho_j|$ clusters parameters with similar optimization behavior.

Hierarchical clustering with complete linkage ensures even the most dissimilar pair within clusters maintains acceptable similarity. The silhouette coefficient [16] evaluates clustering quality across different cluster counts (2 to $K_{\max}$), balancing

intra-cluster cohesion with inter-cluster separation. Optimal cluster count determination employs voting across batches. Each batch independently identifies its optimal count based on maximum silhouette coefficient, with the mode becoming the unified cluster count K^*. Final clustering selection uses Adjusted Rand Index (ARI) [17] validation: the clustering achieving highest total ARI score across all batch pairs ensures maximum consensus.

This automated approach offers significant advantages over block-based clustering. While manual methods cluster parameters by circuit topology regardless of function, ACEMARL discovers functional relationships from data. The complete linkage hierarchical clustering has complexity $O(n \times p^2 + b \times k \times p^2 \log p)$, where n is sample count, p is parameter count, b is batch count, and k is cluster count. This scalability enables application to circuits with hundreds of parameters without expert intervention.

D. GRPO Algorithm and MARL Training

With parameter clusters established, MARL implementation assigns each cluster to a specialized agent sharing observation space $\mathcal{O} = [\mathbf{y}_{\text{ideal}}, \mathbf{y}_{\text{cur}}, \mathbf{P}_{\text{cur}}]$, where $\mathbf{y}_{\text{ideal}}$ are the ideal specifications, $\mathbf{y}_{\text{cur}}$ are the current specifications, and $\mathbf{P}_{\text{cur}}$ are the current circuit parameters. Each agent controls only its assigned cluster $\mathbf{C}_i$ from automated clustering. The framework employs both Group Relative Policy Optimization (GRPO) [18] and Proximal policy optimization (PPO) [19] algorithms. GRPO reduces computational overhead by eliminating the critic network while maintaining comparable performance. Training episodes alternate between BIPOP-CMA-ES samples as initial states and random initialization, balancing exploitation of discovered high-value regions with continued exploration. This integration of evolutionary samples accelerates convergence compared to the state-of-the-art MARL approach.

V. Experiments

A. Experiment Platforms and Cases

All experiments were conducted on a server equipped with AMD EPYC 7C13 processor (2.0 GHz) with 256 GB RAM. Circuit simulations were executed within the Cadence Spectre Simulator using commercial 65 nm and 55 nm CMOS process design kits (PDK).

Three analog circuits were evaluated to validate the superiority of ACEMARL over conventional MARL approaches. The test circuits comprised one amplifier and two LDOs. Fig. 4 illustrated their topologies. Table I summarizes the process, parameter counts, parameter spaces (all possible parameter combinations), and performance specification counts.

All MARL experiments employ a learning rate of 1×10^{-4}. PPO utilizes two identical networks (PolicyNet and ValueNet) for actor-critic training, while GRPO reduces computational overhead by using only PolicyNet. This architectural difference significantly reduces the computational cost.

B. Results and Analysis

The performance of ACEMARL is evaluated through multiple metrics. Maximum reward trajectories indicate the best achievable performance, while mean reward trajectories better demonstrate algorithm robustness. Mean rewards reflect the agents' ability to consistently find proper solutions rather than occasionally discovering optimal ones through random exploration. The confidence intervals of mean rewards quantify training stability. Narrower intervals indicate more reliable convergence behavior, essential for practical deployment where consistent performance matters more than random optimal results.

Fig. 4 illustrates the fundamental difference between manual block-based clustering (SOTA) and ACEMARL automated approach. ACEMARL identifies functional relationship among components that manual methods overlook. For example, ACEMARL correctly isolates power FETs (M7 in LDO #1, M0

TABLE I
SUMMARY OF CIRCUIT PARAMETERS

Topology	Amplifier	LDO #1	LDO #2
Process	65nm	55nm	65nm
Parameter Count	85	179	149
Parameter Space	4.55×10^{80}	1.09×10^{204}	5.71×10^{152}
Specification Count	7	10	14

TABLE II
SUMMARY OF ACEMARL IMPROVEMENT

Topology	Amp.	LDO #1	LDO #2
Max. Reward Acceleration Ratio*	5.0×	3.3×	3.9×
Mean Reward Acceleration Ratio*	1.9×	2.3×	3.4×
Max. Reward Improvement*	18.4%	83.0%	512.6%
Max. FoM Improvement*	5.7%	8.5%	38.5%
Mean Confidence Interval Improvement	31.6%	60.7%	42.1%

* Compare with MAGRPO Curves

TABLE III
SUMMARY OF OPTIMIZATION RESULTS WITH THREE-STAGE AMPLIFIER

Specification	Human Opt.	SOTA Opt.	ACEMARL Opt.
IQ(μA)	2.26	2.57 (-13.7%)	2.25 (0.4%)
Unity Gain Frequency(kHz)	3.52	3.93 (11.7%)	4.02 (14.2%)
Phase Margin(°)	93.34	91.61 (-1.9%)	90.78 (-2.7%)
Gain Margin(dB)	71.60	76.91 (7.4%)	73.12 (2.1%)
Gain(dB)	24.59	30.75 (25.1%)	36.70 (49.3%)
Slew Rate(kV/s)	43.35	50.45 (16.4%)	46.62 (7.5%)
1% Setting Time(μs)	42.40	36.63 (13.6%)	36.82 (13.2%)
FoM	**2.40**	**2.45** (2.1%)	**2.59** (7.9%)

$C_L = 150\text{pF}, \text{VDD} = 1.2V$
Blue: Better than human specs.

TABLE IV
SUMMARY OF OPTIMIZATION RESULTS WITH LDO #1

Specification	Human Opt.	SOTA Opt.	ACEMARL Opt.
IQ(μA)	27.17	26.68 (1.8%)	26.17 (3.7%)
Gain (dB)	88.53	89.94 (1.6%)	96.75 (9.3%)
PM(°)	29.78	28.56 (-4.1%)	55.30 (85.7%)
GBW(kHz)	289.84	315.42 (8.8%)	298.56 (3.0%)
Load Reg.(mV/A)	7.40	5.74 (22.4%)	3.06 (58.7%)
Line Reg.(μV/V)	250.00	103.15 (58.7%)	92.28 (63.1%)
Overshoot(mV)	1.57	2.68 (-70.7%)	0.96 (38.9%)
Undershoot(mV)	5.98	4.88 (18.4%)	4.55 (23.9%)
PSR@100kHz(dB)	33.30	38.44 (15.4%)	40.84 (22.6%)
Startup Shoot (mV)	54.94	93.05 (-69.4%)	52.89 (3.7%)
FoM	**8.94E − 10**	**7.14E − 10** (20.1%)	**6.53E − 10** (27.0%)

[1] ILOAD = 1μA, [2] ILOAD = 20mA
Blue: Better than human specs.

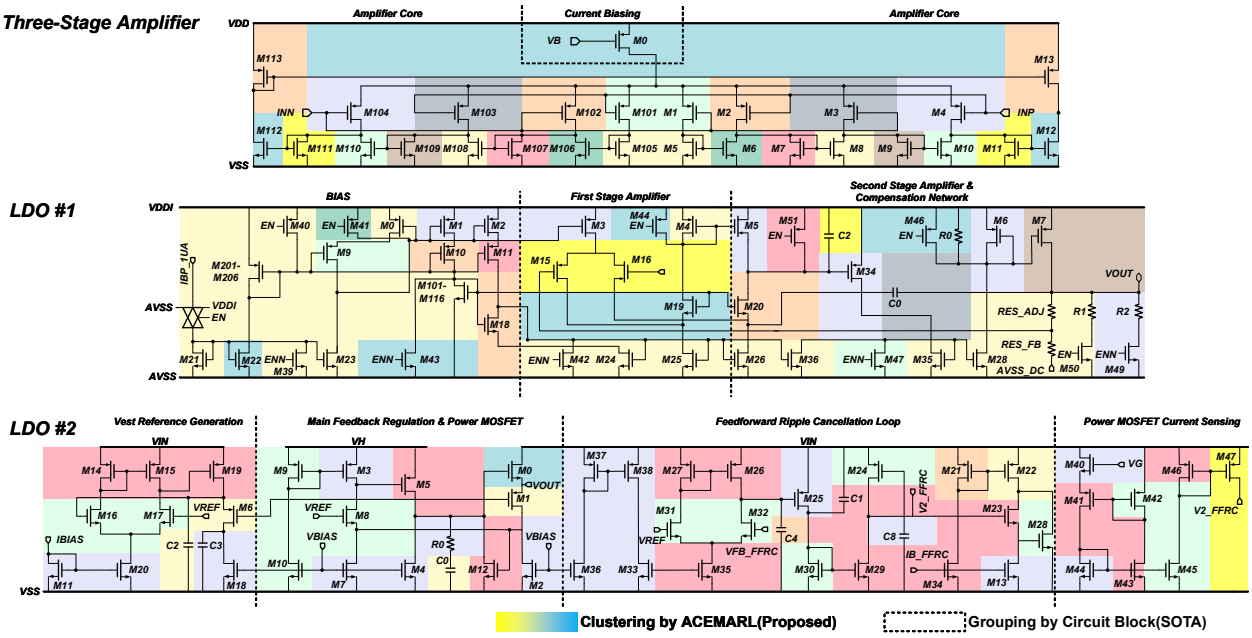

Fig. 4. Transistor-level schematics of test circuits: three-stage amplifier, analog LDO #1, and analog LDO #2. Dashed boxes indicate SOTA manual circuit block clustering with labels. Colored regions show ACEMARL automated clustering results.

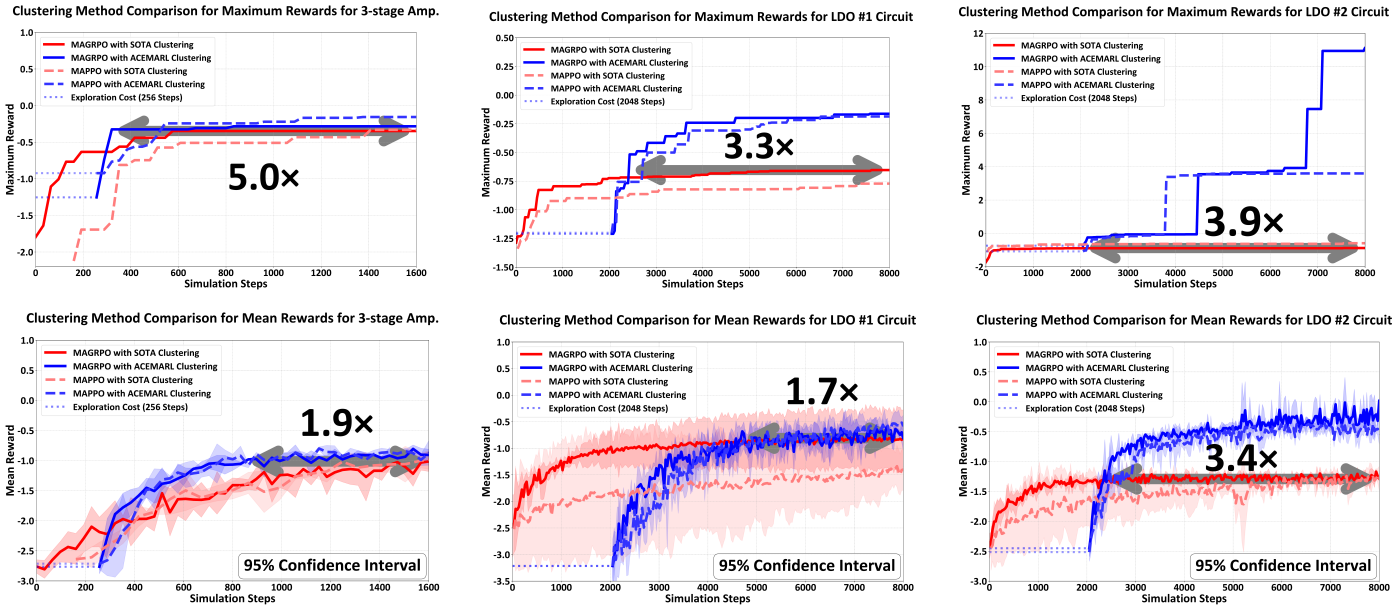

Fig. 5. Training performance comparison between ACEMARL and SOTA clustering methods. Top: maximum rewards; Bottom: mean rewards with 95% confidence intervals. Evaluated on one amplifier and two LDOs using MAPPO and MAGRPO algorithms. ACEMARL consistently achieves superior convergence and stability.

in LDO #2) into separate clusters. These transistors serve as primary power delivery components with the largest chip area, requiring fundamentally different optimization strategies from signal processing transistors. Manual block-based clustering fails to recognize this functional distinction, clustering power FETs with unrelated components.

Fig. 5 demonstrated ACEMARL training advantages across all test circuits. The framework achieved 3.3-5.0× faster maximum reward convergence and 1.9-3.4× faster mean reward convergence as shown in Table II ($\alpha = -5$, $\beta = \gamma = 10$ in Eq. 2). Meanwhile, ACEMARL reduces confidence interval width by 31.6-60.7% for mean reward trajectories, indicating substantially improved training stability. This stability enhancement results from functionally coherent clustering. Agents managing similar components develop consistent policies, reducing variance in the training process.

Circuit-level performance further validates the effectiveness of ACEMARL. Tables III, IV, and V showed consistent improvements across three topologies with ACEMARL. Overall performance is quantified through FoM metrics, providing

TABLE V
SUMMARY OF OPTIMIZATION RESULTS WITH LDO #2

Specification	Human Opt.	SOTA Opt.	ACEMARL Opt.
IQ(μA)	22.11	21.76 (1.6%)	20.11 (9.1%)
Load Reg.(mV/A)	5.31	2.24 (57.8%)	4.51 (15.1%)
PM[1](°)	35.92	46.30 (28.9%)	50.02 (39.2%)
GBW[1](MHz)	36.98	12.05 (-67.4%)	47.70 (29.0%)
PM[2](°)	85.78	65.90 (-23.2%)	87.58 (2.1%)
GBW[2](MHz)	55.30	9.96 (-82.0%)	100.22 (81.2%)
PM[3](°)	85.78	60.60 (-29.4%)	88.64 (3.3%)
GBW[3](MHz)	53.83	5.45 (-89.9%)	97.44 (81.0%)
Overshoot(mV)	176.41	108.32 (38.6%)	132.62 (24.8%)
Undershoot(mV)	227.74	244.40 (-7.3%)	162.60 (28.6%)
PSR[4](dB)	66.05	42.89 (-35.1%)	90.27 (36.7%)
PSR[5](dB)	66.04	42.89 (-35.1%)	90.07 (36.4%)
PSR[6](dB)	36.38	1.24 (-96.6%)	45.33 (24.6%)
PSR[7](dB)	0.45	0.83 (84.4%)	11.79 (2520.0%)
FoM	$\mathbf{4.06E-9}$	$\mathbf{4.29E-9}$ (-5.7%)	$\mathbf{2.64E-9}$ (35.0%)

[1] ILOAD = 200μA, [2] ILOAD = 25mA, [3] ILOAD = 50mA
[4] 10Hz, [5] 1kHz, [6] 10MHz, [7] 100MHz
Blue: Better than human specs.

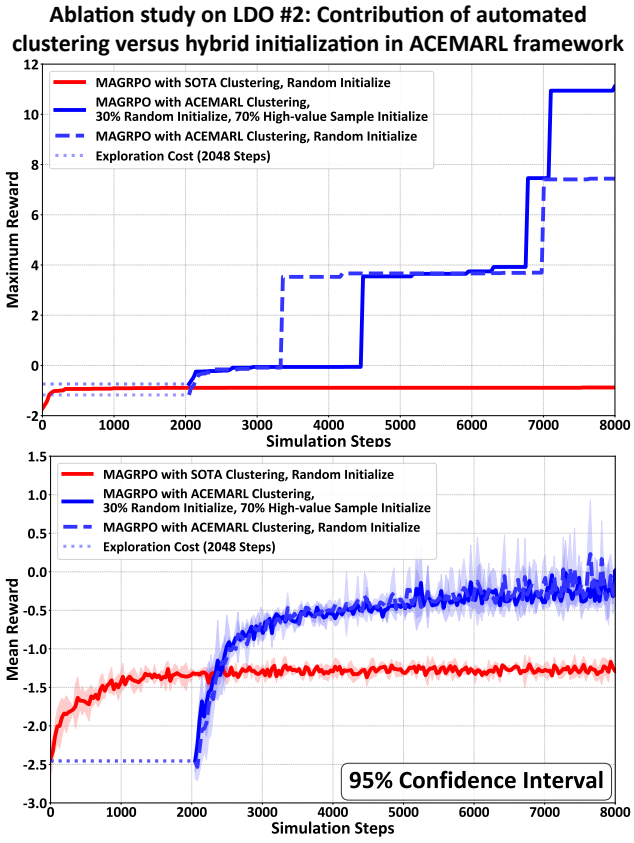

Fig. 6. Ablation study on LDO #2 using MAGRPO: Impact of automated clustering and hybrid initialization. Comparison of (a) SOTA manual clustering, (b) automated clustering only, and (c) complete ACEMARL framework. Mean reward trajectories with 95% confidence intervals are shown.

holistic evaluation beyond individual specifications. For amplifiers, FoM is defined as [20]:

$$\mathbf{FoM}_1 = \frac{\text{SR} \times C_L}{I_Q \times \text{VDD}}, \tag{3}$$

where SR denotes the slew rate, C_L represents the load capacitance, I_Q is the quiescent current, and VDD is the supply voltage. For LDOs, the FoM is expressed as [21]:

$$\mathbf{FoM}_2 = \frac{I_Q C_L \Delta V_{\mathbf{OUT}}}{\Delta I^2_{\mathbf{LOAD}}}, \tag{4}$$

where I_Q is the quiescent current, C_L is the load capacitance, $\Delta V_{\mathbf{OUT}}$ represents the output voltage deviation during load current transients, and $\Delta I_{\mathbf{LOAD}}$ is the load transient current. ACEMARL achieved 5.7-38.5% FoM improvement over SOTA methods.

Table II summarizes the comprehensive advantages of ACEMARL. Consistent improvements in varying circuit topologies demonstrate the robustness of the proposed clustering method.

C. Ablation Study

An ablation study on LDO #2 isolates the contributions of automated clustering and hybrid initialization. Fig. 6 compared three configurations: (1) SOTA block-based clustering with random initialization, (2) automated clustering with random initialization, and (3) complete ACEMARL with hybrid initialization (70% BIPOP-CMA-ES samples, 30% random).

Automated clustering demonstrated dominant impact on performance. While SOTA achieved maximum reward of -0.88, automated clustering alone improved this to 7.44 (69.2% of total improvement) with 3.8× convergence acceleration. Adding hybrid initialization further enhanced maximum reward to 11.13 (contributing remaining 30.8%) with marginal speedup increase to 3.9×. Notably, mean reward convergence remained constant at 3.4× for both automated configurations, confirming that functional clustering, rather than initialization strategy, fundamentally drives ACEMARL's performance improvement. The hybrid initialization primarily enables exploration of superior solutions.

VI. CONCLUSION

This paper presented ACEMARL, a data-driven framework for MARL-based analog circuit sizing. Existing approaches require manual circuit block assignment. ACEMARL discovers functional relationships directly from optimization samples, eliminating the knowledge barrier that limits analog circuit sizing automation. The framework integrates BIPOP-CMA-ES with correlation analysis. This creates a synergistic pipeline where evolutionary exploration provides high-value samples and parameter relationship insights. Experiments validated the approach on diverse circuit topologies with 85-179 parameters. Functional clustering consistently outperforms structural clustering. The framework achieved 3.3-5.0× convergence acceleration for maximum reward trajectories and 5.7-38.5% FoM improvement. Confidence interval width reduced by 31.6-60.7% for mean reward trajectories. This indicates more stable policies where agents specialized on functional clusters outperform those assigned by circuit blocks. These results establish that understanding functional relationships, rather than following circuit topology, is crucial for effective MARL deployment in analog design sizing. While the current framework focuses on single-objective optimization with fixed topologies, the automated clustering methodology opens pathways for multi-objective optimization and topology exploration.

REFERENCES

[1] K. Settaluri, A. Haj-Ali, Q. Huang, K. Hakhamaneshi, and B. Nikolic, "AutoCkt: Deep reinforcement learning of analog circuit designs," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2020, pp. 490–495.

[2] H. Wang, K. Wang, J. Yang, L. Shen, N. Sun, H.-S. Lee, and S. Han, "GCN-RL circuit designer: Transferable transistor sizing with graph neural networks and reinforcement learning," in *ACM/IEEE Design Automation Conference (DAC)*, 2020, pp. 1–6.

[3] J. Zhang, J. Bao, Z. Huang, X. Zeng, and Y. Lu, "Automated design of complex analog circuits with multiagent based reinforcement learning," in *ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6.

[4] M. W. Cohen, M. Aga, and T. Weinberg, "Genetic algorithm software system for analog circuit design," *Procedia CIRP*, vol. 36, pp. 17–22, 2015.

[5] B. Liu, Y. Wang, Z. Yu, L. Liu, M. Li, Z. Wang, J. Lu, and F. V. Fernández, "Analog circuit optimization system based on hybrid evolutionary algorithms," *Integration*, vol. 42, no. 2, pp. 137–148, 2009.

[6] T. Gu, J. Wang, Z. Bi, C. Yan, F. Yang, Y. Qin, T. Cui, and X. Zeng, "tSS-BO: Scalable bayesian optimization for analog circuit sizing via truncated subspace sampling," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2024, pp. 1–6.

[7] S. Zhang, W. Lyu, F. Yang, C. Yan, D. Zhou, and X. Zeng, "Bayesian optimization approach for analog circuit synthesis using neural network," in *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2019, pp. 1463–1468.

[8] W. Lyu, P. Xue, F. Yang, C. Yan, Z. Hong, X. Zeng, and D. Zhou, "An efficient bayesian optimization approach for automated optimization of analog circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 6, pp. 1954–1967, 2018.

[9] Y. Li, Y. Wang, Y. Li, R. Zhou, and Z. Lin, "An artificial neural network assisted optimization system for analog design space exploration," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 39, no. 10, pp. 2640–2653, 2020.

[10] A. F. Budak, P. Bhansali, B. Liu, N. Sun, D. Z. Pan, and C. V. Kashyap, "DNN-Opt: An RL inspired optimization for analog circuit sizing using deep neural networks," in *ACM/IEEE Design Automation Conference (DAC)*, 2021, pp. 1219–1224.

[11] M. Choi, Y. Choi, K. Lee, and S. Kang, "Reinforcement learning-based analog circuit optimizer using gm/ID for sizing," in *ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6.

[12] J. Gao, W. Cao, and X. Zhang, "RoSE: Robust analog circuit parameter optimization with sampling-efficient reinforcement learning," in *ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6.

[13] J. Ruan, X. Hao, D. Li, and H. Mao, "Learning to collaborate by grouping: a consensus-oriented strategy for multi-agent reinforcement learning," *ArXiv*, vol. abs/2307.15530, 2023.

[14] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2020.

[15] N. Hansen, "Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed," in *Annual Conference on Genetic and Evolutionary Computation*, 2009.

[16] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[17] J. M. Santos and M. Embrechts, "On the use of the adjusted rand index as a metric for evaluating supervised classification," in *Artificial Neural Networks – ICANN 2009*, 2009, pp. 175–184.

[18] Z. Shao, P. Wang, Q. Zhu, R. Xu, J. Song, X. Bi, H. Zhang, M. Zhang, Y. K. Li, Y. Wu, and D. Guo, "DeepSeekMath: Pushing the limits of mathematical reasoning in open language models," 2024.

[19] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[20] Z. Yan, P.-I. Mak, M.-K. Law, and R. P. Martins, "A 0.016-mm^2 144-μw three-stage amplifier capable of driving 1-to-15 nf capacitive load with $>$0.95-mhz gbw," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 527–540, 2013.

[21] G. Cai, Y. Lu, C. Zhan, and R. P. Martins, "A fully integrated fvf ldo with enhanced full-spectrum power supply rejection," *IEEE Transactions on Power Electronics*, vol. 36, no. 4, pp. 4326–4337, 2021.