DiversePlace: Diversity-Seeking Curriculum Reinforcement Learning for Macro Placement

Wenrui Zhou* **Jiashun Liu*** **Wenji Fang** **Zhiyao Xie** **Ling Pan†**
HKUST

Abstract

Macro placement is a critical early-stage decision in physical design because macro locations strongly constrain subsequent implementation stages. The task is difficult because these high-impact discrete decisions must be made before downstream quality can be measured accurately, forcing training to rely on fast but imperfect training-time oracles such as half-perimeter wirelength (HPWL). Existing RL-based placers therefore mostly optimize single HPWL-driven solutions, and their long sequential horizons often require heavy search assistance or offline expert data while leaving little room to preserve multiple competitive layouts. We address this gap by making quality-controlled diversity an explicit objective for macro placement. Instead of relying on a single proxy-optimal solution, we maintain a compact archive of HPWL-competitive yet meaningfully different placements for downstream selection. We realize this idea in DiversePlace, a curriculum proximal policy optimization (PPO) framework that combines a quality-gated diversity reward with an expanding placement horizon, enabling from-scratch RL optimization while keeping exploration inside competitive HPWL regions. On the ISPD2005 benchmarks, DiversePlace consistently improves top-5, top-10, and top-20 archive diversity and achieves lower or virtually identical HPWL on all eight designs with better efficiency. Code: `https://github.com/William030422/NIPS2026-DiversePlace`.

1 Introduction

Macro placement is a decisive early step in modern chip design. The macro locations strongly constrain subsequent standard-cell placement, routing, congestion optimization, and timing closure [21, 22]. The critical challenge of macro placement stems from a large legality-constrained combinatorial search space with strong path dependence, where each macro decision reshapes the feasible space for all later decisions. A line of research approached macro placement through optimization and supervised learning. However, gradient-free black-box search [28] is costly on large circuits, analytical relaxation [4, 17] weakens alignment with discrete macro legality and tool-dependent objectives, and supervised learning-driven placers rely on expensive expert layouts [22, 15, 16]. Reinforcement learning (RL) naturally fits macro placement by training a policy to build layouts sequentially on a canvas while maximizing expected long-term rewards, thereby amortizing search over training without relying on expert datasets and achieving promising progress in recent placers.

However, RL-based macro placement is inherently proxy-driven. The final quality of a macro placement is determined only after downstream implementation stages as illustrated in Figure 1, which is unavailable during RL training in this stage. Therefore, existing RL-based placers typically optimize half-perimeter wirelength (HPWL), which is a proxy computed from pin bounding boxes before routed or signoff-oriented metrics are available [17, 22, 14, 15, 10] (detailed in Appendix B).

*Equal contribution.
†Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

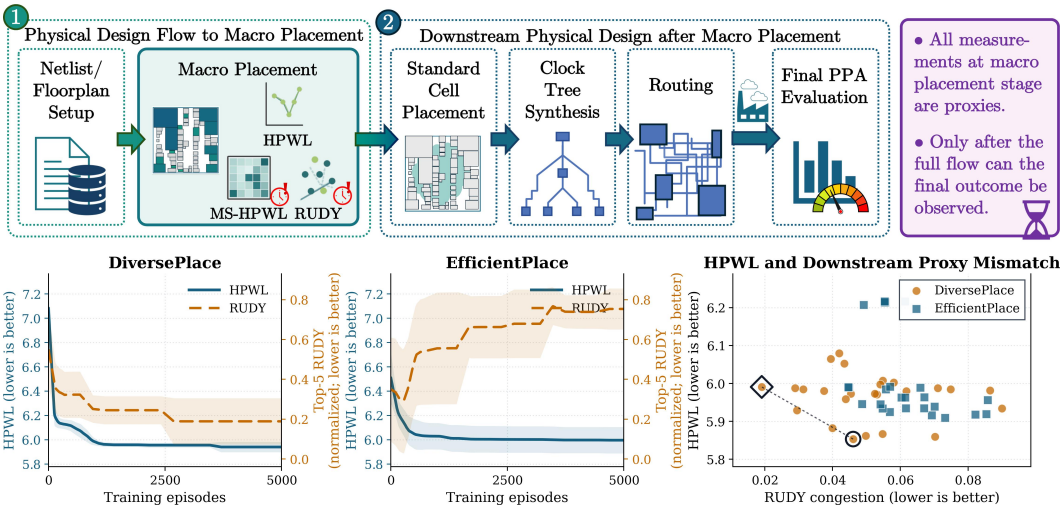


Figure 1: **Proxy-driven macro placement benefits from diverse candidate archives. Top:** Macro placement is optimized before downstream outcomes such as routing and final PPA can be accurately measured, making placement-stage objectives imperfect proxies. **Bottom left/middle:** On *adaptec1*, we evaluate five independent runs (seeds 0–4). Each seed's HPWL-ranked top-5 archive undergoes mixed-size placement and RUDY congestion evaluation. DiversePlace, trained without downstream feedback, yields archives whose best RUDY improves as HPWL decreases, while the similar-mode EfficientPlace archive lacks consistent proxy-to-congestion transfer. **Bottom right:** Best-HPWL and best-RUDY placements do not always coincide. We show that diverse HPWL-competitive archives improve the chance of selecting layouts strong under both HPWL and downstream congestion proxies.

Despite its utility, it is an imperfect training-time oracle for downstream routability, timing, and power, performance, and area (PPA) [29, 5, 12, 1]. The return-optimization nature of RL can therefore drive the policy to overoptimize such an imperfect proxy [9]. Figure 1 illustrates this mismatch on *adaptec1*: HPWL improvement does not necessarily translate into improved downstream congestion measured by rectangular uniform wire density (RUDY, detailed in Appendix D) under a recent RL-based state-of-the-art (SOTA) algorithm [10]. Moreover, sole return optimization in standard RL with imperfect reward can concentrate the policy around one proxy-preferred layout mode [7], while discarding alternative placements that are suboptimal under HPWL but more favorable downstream.

This proxy mismatch exposes an overlooked challenge in RL-based macro placement: the loss of high-quality candidate diversity. Most existing RL-based placers focus on optimizing single-solution proxy quality, implicitly assuming that the best placement under the training reward is also the most useful downstream candidate [14, 15, 10]. However, under an imperfect oracle, this assumption can be fragile. If this mode later underperforms in downstream evaluation, nearby candidates may fail in similar ways, whereas distinct high-quality modes can expose different behaviors across the later physical-design flow [11]. *We therefore argue that an RL placer should return not only one HPWL-minimizing placement, but also a compact archive of high-quality and meaningfully diverse macro placements for downstream selection.*

Motivated by this proxy mismatch, we pursue a diversity-seeking RL formulation for macro placement, where the agent discovers a compact set of HPWL-competitive yet meaningfully different layouts for downstream selection. Realizing this formulation raises two coupled design challenges. **Challenge I. Quality-controlled diversity under an imperfect oracle.** Diversity should be quality-controlled as unconstrained novelty can produce physically irrelevant layouts, whereas overly tight HPWL-based selection can collapse the search to a few proxy-favorable modes and discard other downstream-viable candidates [21, 4, 17, 26]. Most existing learning-based placers overlook this tension by optimizing single-solution proxy quality rather than maintaining a diverse set of HPWL-competitive candidates under an imperfect oracle [22, 5, 14, 15, 10]. **Challenge II. Long-horizon sequential optimization.** Macro placement is a long-horizon decision process in which each trajectory places many macros. Early large-macro decisions reshape the canvas, constrain later feasible regions, and weaken credit assignment across the sequence [14, 10]. This makes diversity-seeking optimization harder, since

the agent must learn to produce competitive placements while exploring distinct high-quality layout modes across a large sequential search space. Thus, the central challenge is not merely to add a diversity term, but to discover distinct layout modes inside a competitive HPWL region while keeping the RL optimization problem learnable.

To realize this formulation, we propose DiversePlace, a novel diversity-seeking curriculum proximal policy optimization (PPO) framework for macro placement [27]. To address quality-controlled diversity, DiversePlace grants diversity rewards only to HPWL-competitive rollouts, and stores their area-weighted macro-displacement descriptors in a quality-diversity (QD) map to encourage new high-quality layout modes [23, 26]. To address long-horizon optimization, DiversePlace trains PPO with an expanding placement curriculum that gradually increases the number of macros controlled by the policy rather than exposing the full horizon from the start [2]. These designs keep sequential optimization learnable while guiding PPO toward multiple HPWL-competitive placement modes.

We evaluate DiversePlace on ISPD2005 macro-placement benchmarks, following prior work [10, 14]. Across eight designs, DiversePlace consistently improves top-5, top-10, and top-20 archive diversity over recent SOTA methods, while achieving lower or virtually identical HPWL at multiple checkpoints with better efficiency. We further evaluate mixed-size placement and congestion, showing that the diverse archive increases the opportunity to select downstream-favorable candidates from a HPWL-ranked top-k set. Ablations validate both the quality-gated diversity objective and the expanding curriculum schedule. The main contributions are summarized as follows.

(1) Quality-controlled diversity for macro placement. We address critical, overlooked diversity loss in RL-based macro placement with a novel QD learning mechanism that encourages multiple diverse, high-quality layouts as competitive candidates for downstream physical-design tools.

(2) Curriculum training for diversity-seeking RL for efficient placement optimization from scratch. We introduce a curriculum PPO framework that progressively expands the placement horizon, enabling efficient short-to-long optimization with faster early HPWL improvement and no loss in final placement quality.

(3) Comprehensive validation and analysis. We extensively analyze optimization strategies, learning designs, and placement diversity, offering practical insights for future method development.

2 Related Work

Optimization and Generative Placement Methods. Classical placement methods span partitioning, stochastic floorplanning, black-box search, and analytical optimization, exposing a long-standing trade-off among feasibility, objective flexibility, and scalability [21]. Stochastic floorplanners with sequence-pair or B*-tree representations and black-box methods such as WireMask-EA directly optimize discrete macro or mixed-size layouts, but their perturb-and-evaluate loops scale poorly on large circuits [13, 24, 3, 28]. Analytical placers such as RePlAce and DREAMPlace scale through continuous optimization of smoothed wirelength and density objectives, but weaken alignment with discrete macro actions, hard legality masks, and non-differentiable tool-dependent rewards [4, 17]. Generative placers, including Flora, GraphPlanner, GAN-Place, and diffusion-based chip placement, learn layout priors from existing or synthetic designs, but depend on data coverage, supervision fidelity, and generalization to unseen netlists, technologies, and constraints [18–20, 16]. These limitations motivate our focus on RL-based macro placement.

RL-Based Methods. RL-based methods are prominent because macro placement naturally admits a sequential MDP formulation [22, 5]. AlphaChip introduced this view, and DeepPR and PRNet extended it toward placement-routing pipelines, but these early methods remained slow and often weak in legality or scalability [22, 5, 6, 14]. MaskPlace improved efficiency with pixel-level masks and dense rewards, and ChiPFormer reduced online cost through offline pretraining, yet both still bear the burden of long-horizon sequential modeling, with the latter additionally depending on offline expert coverage [14, 15]. EfficientPlace further strengthened the quality-runtime trade-off with tree-search-guided PPO, but much of the gain comes from MCTS and sequential freezing rather than from the RL policy alone [10]. More broadly, this line is still largely driven by HPWL or macro-HPWL proxies, while timing and routability are typically evaluated only after placement [22, 14, 15, 10, 28]. This calls for a cleaner RL-native route that better exploits the strengths of policy

learning without tying the agent too tightly to a single proxy reward, thereby leaving more room for stronger overall placement quality across later flow stages.

3 Preliminaries

Macro placement and HPWL. Macro placement assigns a set of hard macros $\mathcal{M} = \{m_1, \ldots, m_T\}$ to legal locations on a chip canvas under boundary and non-overlap constraints. These early macro locations strongly constrain later standard-cell placement, routing, congestion optimization, and timing closure. Since routed wirelength and signoff-oriented objectives are too expensive to evaluate during RL training, learning-based placers commonly optimize half-perimeter wirelength (HPWL) as a fast placement-stage proxy. For a net e with pin set $\mathcal{P}_e$, HPWL is the half-perimeter of the bounding box covering its pins:

$$\mathrm{HPWL} = \sum_{e \in \mathcal{E}} \left(\max_{p \in \mathcal{P}_e} p_x - \min_{p \in \mathcal{P}_e} p_x + \max_{p \in \mathcal{P}_e} p_y - \min_{p \in \mathcal{P}_e} p_y \right).$$

Appendix B illustrates how adding one macro updates the affected HPWL bounding boxes.

Macro placement as sequential decision making. Macro placement can be viewed as a finite-horizon sequential decision-making problem in which a layout is constructed one macro at a time. At step t, the environment contains a partial placement of macros $m_1, \ldots, m_{t-1}$. The state s_t encodes this partial layout together with canvas occupancy, connectivity information, and legality constraints [14]. The action $a_t = (x_t, y_t)$ selects a legal grid location for the current macro m_t; invalid locations are masked out by boundary and non-overlap constraints. After the action is taken, the environment deterministically fixes m_t at the selected location, updates the canvas and legality masks, and proceeds to the next macro. After T decisions, the trajectory yields a complete macro placement. The reward is derived from placement quality, typically through HPWL or its incremental change during placement. We use PPO as the policy optimizer.

4 Method

4.1 Motivation and Overview

DiversePlace is motivated by a training-time oracle mismatch where HPWL serves as a useful training signal but is imperfect as a design target. HPWL is cheap to evaluate online and often correlates with physical-design quality, but downstream stages such as routing, congestion optimization, and timing closure can favor placements that are not the single HPWL winner. Therefore, the HPWL-best placement does not necessarily yield the best downstream outcome, as shown in Figure 2. This motivates a placer that returns multiple HPWL-competitive layouts rather than a single proxy-optimal solution from an imperfect oracle. However, diversity is meaningful only when it remains coupled with quality. A reward that favors difference alone can be satisfied by poor or physically irrelevant placements, whereas HPWL-only optimization tends to collapse to a narrow set of proxy-favorable modes. The goal is therefore to discover multiple distinct layout basins in a competitive HPWL region rather than trade HPWL for superficial spread.

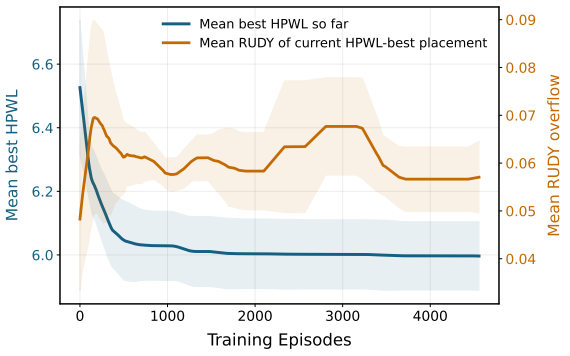


Figure 2: **Best HPWL need not imply the best downstream proxy.** Across five seeds (0–4) on Adaptec1, the current SOTA algorithm has non-monotonic mean RUDY overflow at its current HPWL-best placement.

DiversePlace addresses this problem with a quality-gated diversity objective and a curriculum that preserves optimization quality over long horizons. First, in each curriculum window, PPO runs a

warmup phase under the base placement reward to establish a local HPWL baseline and collect high-quality reference placements. Second, the diversity phase grants a capped terminal bonus only to rollouts whose final HPWL remains within a small tolerance of that baseline, preventing novelty search from drifting into low-quality regions. For each quality-eligible rollout, the bonus is computed from area-weighted, size-normalized macro displacement relative to the nearest high-quality reference and inserted into a two-dimensional quality-diversity (QD) map, whose cells retain the best-HPWL elite for each movement pattern. Third, to keep this objective stable as the horizon grows, the next window expands the PPO-controlled horizon, reuses only a short prefix bridge selected from top placements in the previous window, and completes the remaining suffix greedily so every rollout is still evaluated as a full placement. This design lets PPO explore alternative layout basins while remaining anchored to placements that are still competitive under the imperfect reward signal.

4.2 Diversity-Seeking Reward

Quality-gated references. DiversePlace applies diversity to completed placements rather than partial rollouts. In each episode, PPO places the macros currently exposed by the curriculum, the remaining suffix is completed greedily, and the resulting full placement is evaluated once at the terminal transition. Diversity is meaningful only among competitive placements. Therefore, at each curriculum window k, DiversePlace first runs a warmup phase under the base placement reward and records the best warmup placement $P_k^\star$. It then forms a reference set $\mathcal{R}_k$ from warmup placements whose HPWL lies within a small tolerance of this best value:

$$\mathcal{R}_k = \{P : \mathrm{HPWL}(P) \le (1+\epsilon_q)\mathrm{HPWL}(P_k^\star)\}. \tag{1}$$

During the subsequent diversity phase, a rollout outside this HPWL band receives no diversity bonus, even if it differs substantially from previous placements. This gate restricts novelty search to a local region of placements that remain competitive in proxy quality.

Normalized displacement. Given a completed placement P and a reference placement $R \in \mathcal{R}_k$, DiversePlace measures novelty from the movement of macros in the current diversity scope $\mathcal{S}_k$. For macro m_i with placement coordinates (x_i, y_i) in P, reference coordinates (x_i^R, y_i^R) in R, and size (w_i, h_i), the normalized displacement is

$$d_i(P,R) = \frac{\sqrt{(x_i - x_i^R)^2 + (y_i - y_i^R)^2}}{\sqrt{w_i^2 + h_i^2}}. \tag{2}$$

We aggregate these displacements with area weights $\omega_i = w_i h_i / \sum_j w_j h_j$:

$$D(P,R) = \sum_{i \in \mathcal{S}_k} \omega_i d_i(P,R). \tag{3}$$

Diagonal normalization makes movement comparable across macro sizes, while area weighting gives larger macros more influence because they shape the global placement basin more strongly.

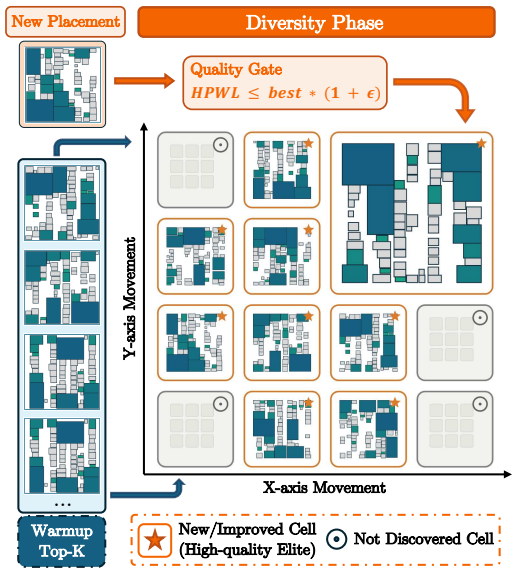


Figure 3: **Quality-gated diversity.** HPWL-gated candidate placements enter a QD map parameterized by area-weighted, size-normalized displacement from the nearest high-quality reference.

QD descriptor and elite map. DiversePlace compares each placement with its nearest high-quality reference,

$$R^\star(P) = \arg\min_{R \in \mathcal{R}_k} D(P,R), \tag{4}$$

so novelty is measured relative to the most similar competitive mode rather than to an arbitrary reference. From $R^\star(P)$, DiversePlace builds a two-dimensional descriptor that separates movement along the two canvas axes:

$$b(P) = \left(g\left(\sum_{i \in \mathcal{S}_k} \omega_i \frac{|x_i - x_i^{R^\star(P)}|}{\sqrt{w_i^2 + h_i^2}} \right), g\left(\sum_{i \in \mathcal{S}_k} \omega_i \frac{|y_i - y_i^{R^\star(P)}|}{\sqrt{w_i^2 + h_i^2}} \right) \right), \tag{5}$$

where $g(z) = z/(1+z)$ maps each coordinate to a bounded range. The descriptor is inserted into a two-dimensional QD map whose bin edges are fit within each window from the warmup descriptor distribution. Each cell stores the lowest-HPWL elite discovered for that movement pattern. Empty cells represent unexplored layout modes, whereas occupied cells retain the best placement found in that region of the descriptor space.

Terminal diversity bonus. The QD map changes training through the terminal PPO reward. If placement P passes the quality gate, its raw diversity bonus is

$$r_{\mathrm{div}}(P) = \lambda D(P, R^{\star}(P))\, \eta(c(P)), \tag{6}$$

where $c(P)$ is the QD cell of P and

$$\eta(c) = \begin{cases} 1, & \text{if } c \text{ is empty or } P \text{ improves the cell elite,} \\ (1+n_c)^{-1/2}, & \text{otherwise.} \end{cases} \tag{7}$$

with n_c the visit count of cell c. The final bonus is capped relative to the terminal placement reward:

$$\tilde{r}_{\mathrm{div}}(P) = \min\left(r_{\mathrm{div}}(P),\, \beta |r_{\mathrm{term}}|\right). \tag{8}$$

The terminal PPO transition stores $r_{\mathrm{term}} + \tilde{r}_{\mathrm{div}}$, while intermediate rewards remain unchanged. Therefore, the QD map is not merely an evaluation archive. It directly biases PPO toward quality-eligible placements that enter new cells or improve existing elites.

4.3 Curriculum over Placement Horizons

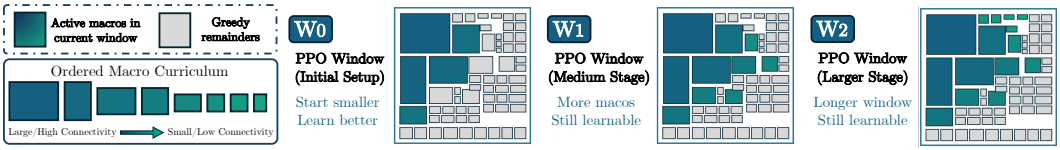


Figure 4: **Curriculum over placement horizons.** DiversePlace starts from a short PPO-controlled macro window and progressively expands it, while greedily completing the remaining macros so every rollout is evaluated as a full placement.

Why curriculum is needed. The diversity reward is useful only if PPO can first reach a competitive HPWL region. This is difficult when the policy must solve the full macro-placement horizon from the beginning, because early macro decisions determine the feasible basin for many later placements and make credit assignment increasingly weak. DiversePlace therefore uses curriculum learning to make quality-gated diversity compatible with long-horizon RL optimization.

Expanding horizon. Let T be the total number of macros and let H_k be the number of macros controlled by PPO in curriculum window k. DiversePlace starts from a small horizon H_0 and expands it by a fixed increment:

$$H_k = \min(T, H_0 + k\Delta H). \tag{9}$$

Within window k, PPO places macros $m_1, \ldots, m_{H_k}$ using the visual state, legal-position mask, and wire mask. The remaining macros $m_{H_k+1}, \ldots, m_T$ are then placed greedily so that every episode still produces a complete placement and a full HPWL value. After each window, the horizon expands and the same actor-critic networks continue training. This schedule reduces the initial credit-assignment burden while preserving the full-placement objective at every episode.

Window schedule. Each curriculum window has two phases. The warmup phase optimizes the base placement reward and establishes a local HPWL baseline together with the reference set for the current horizon. The diversity phase then searches for alternative placements within that local quality region using the QD bonus defined above. Because the gate and the references are recomputed at every horizon, the diversity signal stays aligned with the difficulty of the current window instead of relying on stale placements from earlier, shorter horizons.

Prefix bridge. At horizon boundaries, DiversePlace ranks completed placements from the previous window by HPWL and retains a small top-K prefix pool. The next window reuses only the first few macro locations from one of these placements as a short prefix bridge. This transfer stabilizes early high-impact decisions that largely determine the feasible basin for later macros, while leaving most of the newly exposed horizon free for PPO refinement. The transferred segment is deliberately small (e.g., two macros), allowing the policy to still reshape the placement as the horizon expands. Together, the expanding horizon, per-window warmup-then-diversity schedule, and short prefix bridge keep PPO in a competitive HPWL region with enough freedom to discover multiple layout modes.

5 Experiments

5.1 Benchmarks and Settings

We evaluate DiversePlace against one classical reference baseline, simulated annealing (SA) [13], and five RL-based placers including GraphPlace [22], DeepPR [5], MaskPlace [14], ChiPFormer [15], and EfficientPlace [10]. All runs in this work, including the EfficientPlace reruns and all DiversePlace experiments, are conducted on a single server with two Intel Xeon Platinum 8375C CPUs and eight NVIDIA RTX 3090 GPUs, each with 24 GB of memory; each seed uses one GPU. Consistent with recent macro-placement studies [14, 15, 10], we evaluate all methods on the widely used ISPD2005 benchmark suite [25], which contains the eight instances *adaptec1–4* and *bigblue1–4*.

5.2 Evaluation Metrics

We use HPWL, reported in units of 10^5, as the primary placement-quality metric; lower values are better. For comparisons with the matching budget, we report the cumulative-best HPWL after global episodes of 1K, 2K, and 10K. To assess whether a method preserves multiple competitive placement modes, we also analyze the global top-k solution archive saved during training. For archive diversity, we reuse the area-weighted, size-normalized macro-displacement defined in Section 4.2, but apply it pairwise between complete placements over their shared comparable macro set. For an archive $\{P_i\}_{i=1}^{k}$, the diversity score is the mean pairwise distance

$$D_k = \frac{2}{k(k-1)} \sum_{i<j} d(P_i, P_j). \tag{10}$$

This pairwise-archive view follows standard set-diversity practice in quality-diversity methods [23, 26]. Thus, D_k is a unitless measure of geometric variation among high-quality placements; higher values indicate a broader set of distinct layouts.

5.3 Results and Analysis

HPWL Results. Table 1 presents the overall HPWL comparison against prior macro-placement methods. Under a controlled protocol, DiversePlace is faster than EfficientPlace on seven of the eight benchmarks at 1K episodes. It also delivers lower or essentially comparable HPWL on most datasets, achieving the best 10K HPWL on seven benchmarks.

The presented checkpoints in Table 1 and full trajectories in Appendix Figure C1 show that DiversePlace's advantage emerges early rather than only during late-stage training. We argue that placement quality is largely determined by the early basin reached by the policy: PPO mainly steers search toward stronger initial basins, after which greedy placement induced by WireMask and DreamPlace refinement can recover much of the remaining quality [14]. By 1K and 2K episodes, DiversePlace already achieves lower HPWL than EfficientPlace on six of eight benchmarks, increasing this count to seven by 10K episodes. Moreover, by incorporating curriculum, we further show that this early quality does not require a larger wall-clock budget on most benchmarks: DiversePlace uses $0.73\times$

Table 1: HPWL values ($\times 10^5$) obtained by compared macro-placement methods on eight ISPD2005 circuits [25]. Results for prior methods are taken from the corresponding reported sources [22, 5, 14, 15, 10], whereas EfficientPlace and DiversePlace are evaluated using a single predetermined batch of five independent runs with seeds 0–4, all of which are included when reporting means and sample standard deviations. Numbers after method names give the global-episode counts. For the 1K and 2K rows, runtime in hours is reported below HPWL in parentheses. Lower HPWL and runtime are better; **bold** and underlining mark the best and second-best HPWL results.

Method	adaptec1	adaptec2	adaptec3	adaptec4	bigblue1	bigblue2	bigblue3	bigblue4
SA (1M)	19.24±1.42	102.84±5.83	140.34±7.48	148.57±23.08	5.92±0.98	12.09±0.42	270.46±43.76	248.38±12.25
GraphPlace (50K)	30.01±2.98	351.71±38.20	358.18±13.95	151.42±9.72	10.58±1.29	14.78±0.95	357.48±47.83	440.70±15.95
DeepPR (3K)	19.91±2.13	203.51±6.27	347.16±4.32	311.86±56.74	23.33±3.65	11.38±0.20	430.48±12.18	433.90±5.26
MaskPlace (3K)	7.62±0.67	75.16±4.97	100.24±13.54	87.99±3.25	3.04±0.06	5.75±0.11	90.04±4.83	103.26±2.69
ChiPFormer (2K)	6.62±0.05	67.10±5.46	76.70±1.15	68.80±1.59	2.95±0.04	5.44±0.10	72.92±2.56	102.84±0.15
EfficientPlace (1K)	6.03±0.12 (0.82h)	47.19±3.50 (0.87h)	57.42±1.53 (1.72h)	61.32±1.30 (4.25h)	2.29±0.03 (0.91h)	4.75±0.13 (0.37h)	74.96±21.93 (3.20h)	89.88±1.95 (2.29h)
DiversePlace (1K)	6.01±0.09 (0.45h)	46.81±0.68 (0.63h)	56.93±0.74 (1.24h)	59.64±0.84 (2.88h)	2.31±0.02 (0.95h)	4.78±0.12 (0.19h)	73.10±11.92 (2.61h)	88.60±2.28 (1.79h)
EfficientPlace (2K)	6.00±0.12 (1.47h)	46.40±3.54 (1.65h)	57.29±1.68 (3.33h)	60.57±1.44 (8.50h)	2.29±0.03 (1.70h)	4.69±0.13 (0.69h)	71.43±23.44 (6.23h)	88.33±1.78 (4.51h)
DiversePlace (2K)	5.99±0.07 (0.91h)	44.59±1.15 (1.28h)	55.48±0.54 (2.51h)	58.66±0.89 (5.64h)	2.29±0.02 (2.45h)	4.70±0.04 (0.41h)	61.81±15.64 (5.19h)	85.40±2.53 (3.50h)
EfficientPlace (10K)	6.00±0.12	46.12±3.45	56.46±1.52	59.19±1.46	**2.28±0.02**	4.57±0.08	68.65±22.07	85.71±1.49
DiversePlace (10K)	**5.91±0.05**	**42.44±0.87**	**55.25±0.66**	**56.40±0.92**	2.29±0.02	**4.56±0.05**	**59.48±14.62**	**82.91±2.70**

Table 2: Top-K archive quality and diversity comparison. Each cell reports DiversePlace / EfficientPlace, with the winning value highlighted in **bold**. HPWL rows report archive quality, where lower is better. D_k^{full} reports mean pairwise normalized macro-displacement over complete placements. D_k^{PPO} removes the WireMask greedy tail and evaluates only the PPO-controlled macro range.

Metric	k	adaptec1	adaptec2	adaptec3	adaptec4	bigblue1	bigblue2	bigblue3	bigblue4
HPWL	5	**5.96** / 6.00	**42.77** / 46.24	**55.82** / 56.87	**57.16** / 59.88	2.29 / 2.29	4.61 / 4.61	**60.92** / 69.75	**85.05** / 86.79
HPWL	10	**5.98** / 6.00	**42.92** / 46.33	**56.01** / 57.05	**57.58** / 60.21	2.29 / 2.29	4.63 / **4.62**	**61.77** / 70.32	**86.03** / 87.28
HPWL	20	**6.00** / 6.01	**43.08** / 46.42	**56.25** / 57.24	**58.00** / 60.55	2.30 / **2.29**	4.66 / **4.64**	**62.75** / 70.98	**86.89** / 87.90
D_k^{full}	5	**3.98** / 2.13	**1.83** / 1.48	**4.20** / 3.41	**10.89** / 8.56	**3.80** / 2.79	**2.28** / 0.62	**4.57** / 3.37	**6.13** / 4.42
D_k^{full}	10	**3.89** / 2.17	**1.77** / 1.51	**4.28** / 3.62	**11.42** / 8.75	**3.82** / 2.76	**2.20** / 0.62	**4.66** / 3.38	**6.24** / 4.27
D_k^{full}	20	**3.86** / 2.17	**1.75** / 1.53	**4.34** / 3.68	**11.69** / 8.82	**3.84** / 2.71	**2.49** / 0.67	**4.65** / 3.47	**6.26** / 4.28
D_k^{PPO}	5	**1.87** / 0.01	**0.34** / 0.01	**0.36** / 0.02	**0.68** / 0.19	**1.10** / 0.15	**2.32** / 0.13	**0.86** / 0.10	**0.70** / 0.40
D_k^{PPO}	10	**1.80** / 0.02	**0.27** / 0.01	**0.40** / 0.11	**0.71** / 0.19	**1.12** / 0.11	**2.22** / 0.10	**0.85** / 0.10	**0.69** / 0.30
D_k^{PPO}	20	**1.81** / 0.01	**0.26** / 0.01	**0.38** / 0.12	**0.76** / 0.24	**1.13** / 0.06	**2.54** / 0.13	**0.84** / 0.14	**0.72** / 0.29

the 1K runtime and $0.81\times$ the 2K runtime of EfficientPlace on average. These results indicate that our paradigm improves early solution quality while preserving practical training efficiency.

Diversity Results. Table 2 reports two diversity views. D_k^{full} is the original archive diversity over complete placements, while D_k^{PPO} removes the WireMask greedy tail and measures only the macros still inside PPO control [28]. DiversePlace is higher than EfficientPlace for every benchmark and every archive size in both views. In addition, Figure 5 visualizes the archived placements on *adaptec4*, providing a qualitative view of the layout modes captured by the diversity mechanism.

Downstream Evaluation. We further evaluate whether the archive produced by DiversePlace transfers to downstream physical-design proxies beyond macro-placement HPWL. Specifically, we complete standard-cell placement with DREAMPlace [17] and evaluate routability using RUDY congestion [29], with detailed protocols and results reported in Appendix D. Although DiversePlace is trained primarily with HPWL, its diverse top-k archive provides a more favorable candidate pool for downstream selection, creating additional opportunities to identify HPWL-competitive macro placements that align better with mixed-size placement quality or congestion profiles. Moreover, full placement-and-routing evaluation on *ariane133* demonstrates higher candidate yield and stronger post-route timing–wirelength results across three training seeds (Appendix G). These results support the role of quality-controlled diversity as a practical interface between proxy-driven macro placement and later physical-design stages, rather than merely as an exploration mechanism.

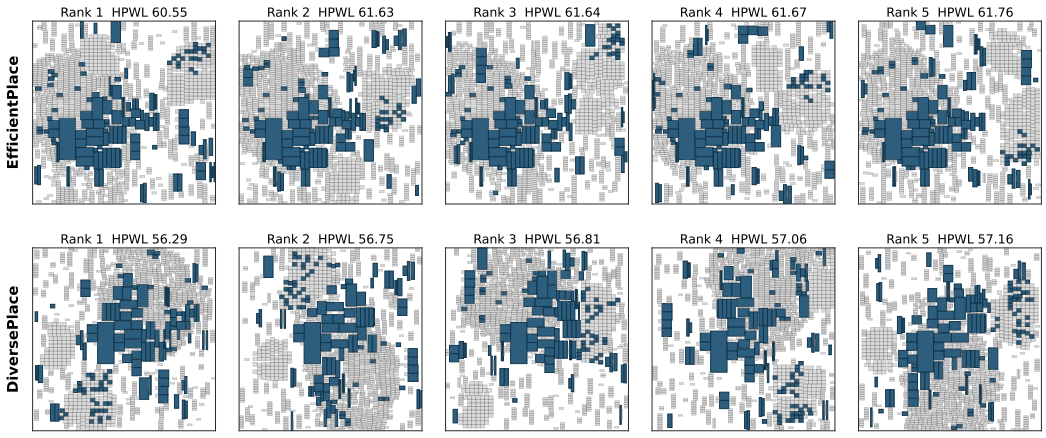


Figure 5: **Top-5 archive placements on *adaptec4*.** We compare EfficientPlace and DiversePlace top-5 archive candidates from selected independent runs. The largest 128 macros by area are highlighted in **blue** to make the largest macros' diverse placement modes more prominent. A corresponding mixed-size visualization of these archive placements is provided in Appendix F.

5.4 Ablation Studies

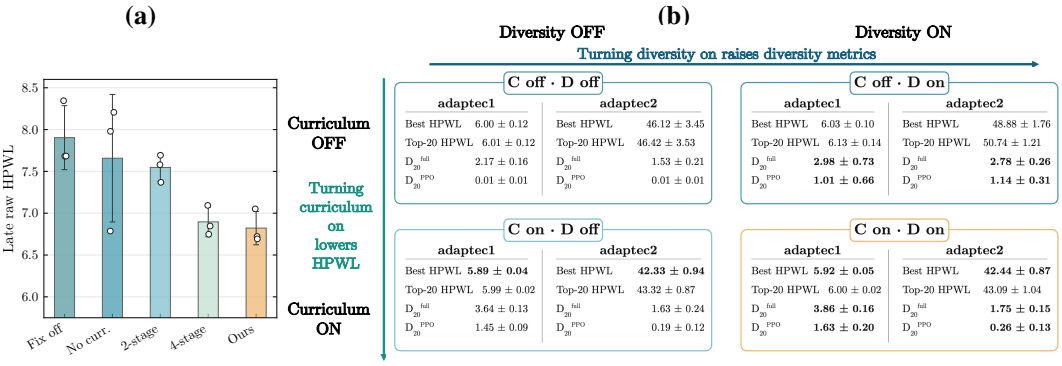


Figure 6: **(a)** Late-stage raw HPWL on *adaptec1*, episodes 6K–10K. **(b)** Unified 2×2 curriculum–diversity ablation on *adaptec1* and *adaptec2*. Lower HPWL and higher D_{20} are better.

Figure 6(a) reports late-stage raw HPWL on *adaptec1* over the final 4K global episodes, after early transient exploration has largely subsided. Removing prefix transfer gives the worst HPWL, indicating that inherited prefixes are not merely warm starts but structural priors that anchor previously optimized large macros and shape subsequent layout formation. Among the tested curriculum variants, a finer staged expansion, corresponding to a smaller ΔH in Section 4.3, yields lower final-window HPWL, suggesting that a reasonably gradual curriculum can provide a smoother optimization path for late-stage placement search. DiversePlace therefore achieves the lowest late-stage HPWL by combining a small prefix-induced stability with a finer curriculum path, avoiding the less reliable optimization induced by exposing PPO to the full macro horizon at once.

Figure 6(b) presents a two-by-two ablation of the curriculum and diversity components in DiversePlace. Across both *adaptec1* and *adaptec2*, diversity alone substantially expands the archive but does not reliably improve HPWL, indicating that unconstrained exploration can discover different placements without providing an effective optimization path. By contrast, curriculum accounts for the main HPWL gains, making the long-horizon placement problem easier to optimize, but its archive remains less diverse than that of the full method. Adding diversity on top of curriculum consistently increases both full-placement and PPO-region diversity while preserving the HPWL advantage obtained by curriculum. Overall, the ablation reveals a clear division of labor: curriculum

makes high-quality placement search tractable, whereas quality-guided diversity broadens the set of competitive layouts discovered within that optimized search process. Appendix E provides a more detailed analysis of key parameters.

6 Conclusion and Limitations

Conclusion. We presented DiversePlace, a diversity-seeking curriculum PPO framework for macro placement that searches for alternative layout modes inside a competitive HPWL region. By coupling a quality-gated QD objective with an expanding placement-horizon curriculum, DiversePlace improves archive diversity on all eight ISPD2005 benchmarks while also delivering stronger early efficiency and lower or closely comparable 10K-checkpoint HPWL than a recent SOTA RL baseline on all designs. These results suggest quality-controlled diversity as a useful design principle for RL-based physical design when downstream quality matters but is not directly measurable during training.

Limitations. DiversePlace targets early-stage macro placement under an HPWL-anchored quality gate, matching the practical setting where accurate routing or timing feedback remains too expensive to query online at every rollout. This scope lets the method preserve a fast and stable RL loop while already exposing multiple competitive layout basins, and it leaves open the integration of richer downstream surrogates when they can be incorporated at comparable cost.

Acknowledgments

The work of Wenrui Zhou, Jiashun Liu, and Ling Pan was supported in part by National Natural Science Foundation of China 62406266. The work of Wenji Fang and Zhiyao Xie was supported in part by (RGC) CRF-YCRG C6003-24Y.

References

[1] Anthony Agnesina, Puranjay Rajvanshi, Tian Yang, Geraldo Pradipta, Austin Jiao, Ben Keller, Brucek Khailany, and Haoxing Ren. AutoDMP: Automated DREAMPlace-based macro placement. In *Proceedings of the 2023 International Symposium on Physical Design*, ISPD '23, pages 149–157. Association for Computing Machinery, 2023. doi: 10.1145/3569052.3578923.

[2] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th International Conference on Machine Learning*, pages 41–48, 2009. doi: 10.1145/1553374.1553380.

[3] Yun-Chih Chang, Yao-Wen Chang, Guang-Ming Wu, and Shu-Wei Wu. B*-trees: A new representation for non-slicing floorplans. In *Proceedings of the 37th Design Automation Conference*, pages 458–463, 2000. doi: 10.1145/337292.337541.

[4] Chung-Kuan Cheng, Andrew B. Kahng, Ilgweon Kang, and Lutong Wang. RePlAce: Advancing solution quality and routability validation in global placement. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 38(9):1717–1730, 2019. doi: 10.1109/TCAD.2018.2859220.

[5] Ruoyu Cheng and Junchi Yan. On joint learning for solving placement and routing in chip design. In *Advances in Neural Information Processing Systems*, volume 34, pages 16508–16519, 2021.

[6] Ruoyu Cheng, Xianglong Lyu, Yang Li, Junjie Ye, Jianye Hao, and Junchi Yan. The policy-gradient placement and generative routing neural networks for chip design. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

[7] Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=SJx63jRqFm`.

[8] Chengrui Gao, Yunqi Shi, Ke Xue, Ruo-Tong Chen, Siyuan Xu, Mingxuan Yuan, Chao Qian, and Zhi-Hua Zhou. Expertise can be helpful for reinforcement learning-based macro placement. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust, editors, *International Conference on Learning Representations*, volume 2026, pages 125611–125634, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/cbc1ad2066f0afebbcea930c5688fc1f-Paper-Conference.pdf.

[9] Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 10835–10866. PMLR, 2023.

[10] Zijie Geng, Jie Wang, Ziyan Liu, Siyuan Xu, Zhentao Tang, Mingxuan Yuan, Jianye Hao, Yongdong Zhang, and Feng Wu. Reinforcement learning within tree search for fast macro placement. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 15402–15417, 2024.

[11] Moksh Jain, Emmanuel Bengio, Alex Hernandez-Garcia, Jarrid Rector-Brooks, Bonaventure F. P. Dossou, Chanakya Ajit Ekbote, Jie Fu, Tianyu Zhang, Michael Kilgour, Dinghuai Zhang, Lena Simine, Payel Das, and Yoshua Bengio. Biological sequence design with GFlowNets. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9786–9801. PMLR, 2022. URL https://proceedings.mlr.press/v162/jain22a.html.

[12] Andrew B. Kahng. Advancing placement. In *Proceedings of the 2021 International Symposium on Physical Design*, ISPD '21, pages 15–22. Association for Computing Machinery, 2021. doi: 10.1145/3439706.3446884.

[13] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983. doi: 10.1126/science.220.4598.671.

[14] Yao Lai, Yao Mu, and Ping Luo. Maskplace: Fast chip placement via reinforced visual representation learning. *Advances in Neural Information Processing Systems*, 35:24019–24030, 2022.

[15] Yao Lai, Jinxin Liu, Zhentao Tang, Bin Wang, Jianye Hao, and Ping Luo. ChiPFormer: Transferable chip placement via offline decision transformer. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 18346–18364, 2023.

[16] Vint Lee, Minh Nguyen, Leena Elzeiny, Chun Deng, Pieter Abbeel, and John Wawrzynek. Chip placement with diffusion models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 33499–33514, 2025. URL https://proceedings.mlr.press/v267/lee25y.html.

[17] Yibo Lin, Zixuan Jiang, Jiaqi Gu, Wuxi Li, Shounak Dhar, Haoxing Ren, Brucek Khailany, and David Z. Pan. DREAMPlace: Deep learning toolkit-enabled GPU acceleration for modern VLSI placement. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 40(4):748–761, 2021. doi: 10.1109/TCAD.2020.3003843.

[18] Yiting Liu, Ziyi Ju, Zhengming Li, Mingzhi Dong, Hai Zhou, Jia Wang, Fan Yang, Xuan Zeng, and Li Shang. Floorplanning with graph attention. In *Proceedings of the 59th ACM/IEEE Design Automation Conference*, pages 1303–1308, 2022. doi: 10.1145/3489517.3530484.

[19] Yiting Liu, Ziyi Ju, Zhengming Li, Mingzhi Dong, Hai Zhou, Jia Wang, Fan Yang, Xuan Zeng, and Shang Li. GraphPlanner: Floorplanning with graph neural network. *ACM Transactions on Design Automation of Electronic Systems*, 28(2):21:1–21:24, 2023. doi: 10.1145/3555804.

[20] Yi-Chen Lu, Haoxing Ren, Hao-Hsiang Hsiao, and Sung Kyu Lim. GAN-Place: Advancing open source placers to commercial-quality using generative adversarial networks and transfer learning. *ACM Transactions on Design Automation of Electronic Systems*, 29(2):1–17, 2024. doi: 10.1145/3636461.

[21] Igor L. Markov, Jin Hu, and Myung-Chul Kim. Progress and challenges in VLSI placement research. *Proceedings of the IEEE*, 103(11):1985–2003, 2015. doi: 10.1109/JPROC.2015.2478963.

[22] Azalia Mirhoseini, Anna Goldie, Mustafa Yazgan, Joe Wenjie Jiang, Ebrahim Songhori, Shen Wang, Young-Joon Lee, Eric Johnson, Omkar Pathak, Azade Nova, Jiwoo Pak, Andy Tong, Kavya Srinivasa, William Hang, Emre Tuncer, Quoc V. Le, James Laudon, Richard Ho, Roger Carpenter, and Jeff Dean. A graph placement methodology for fast chip design. *Nature*, 594 (7862):207–212, 2021. doi: 10.1038/s41586-021-03544-w.

[23] Jean-Baptiste Mouret and Jeff Clune. Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909*, 2015. doi: 10.48550/arXiv.1504.04909. URL https://arxiv.org/abs/1504.04909.

[24] Hiroshi Murata, Kunihiro Fujiyoshi, Shinji Nakatake, and Yoshiyuki Kajitani. Vlsi module placement based on rectangle-packing by the sequence-pair. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 15(12):1518–1524, 1996. doi: 10.1109/43.552084.

[25] Gi-Joon Nam, Charles J. Alpert, Paul Villarrubia, Bruce Winter, and Mehmet Can Yildiz. The ISPD2005 placement contest and benchmark suite. In *Proceedings of the 2005 International Symposium on Physical Design*, pages 216–220, 2005. doi: 10.1145/1055137.1055182.

[26] Justin K. Pugh, Lisa B. Soros, and Kenneth O. Stanley. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI*, 3:40, 2016. doi: 10.3389/frobt.2016.00040. URL https://www.frontiersin.org/articles/10.3389/frobt.2016.00040/full.

[27] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. doi: 10.48550/arXiv.1707.06347. URL https://arxiv.org/abs/1707.06347.

[28] Yunqi Shi, Ke Xue, Song Lei, and Chao Qian. Macro placement by wire-mask-guided black-box optimization. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

[29] Peter Spindler and Frank M. Johannes. Fast and accurate routing demand estimation for efficient routability-driven placement. In *Design, Automation and Test in Europe Conference and Exhibition*, pages 1–6. IEEE, 2007. doi: 10.1109/DATE.2007.364463.

[30] Ke Xue, Ruo-Tong Chen, Xi Lin, Yunqi Shi, Shixiong Kai, Siyuan Xu, and Chao Qian. Reinforcement learning policy as macro regulator rather than macro placer. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 140565–140588. Curran Associates, Inc., 2024. doi: 10.52202/079017-4462. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/fe224a60b878e79d5b3d79d7f113f76b-Paper-Conference.pdf.

A Dataset Description

We evaluate on the eight public ISPD2005 placement-contest circuits, *adaptec1–4* and *bigblue1–4* [25] following prior work [14, 10]. We treat each circuit as a separate placement instance with its own canvas, fixed terminals, macro geometry, standard-cell population, netlist connectivity, and utilization profile. Table A1 summarizes the public circuit statistics reported for the ISPD2005 instances used in our experiments.

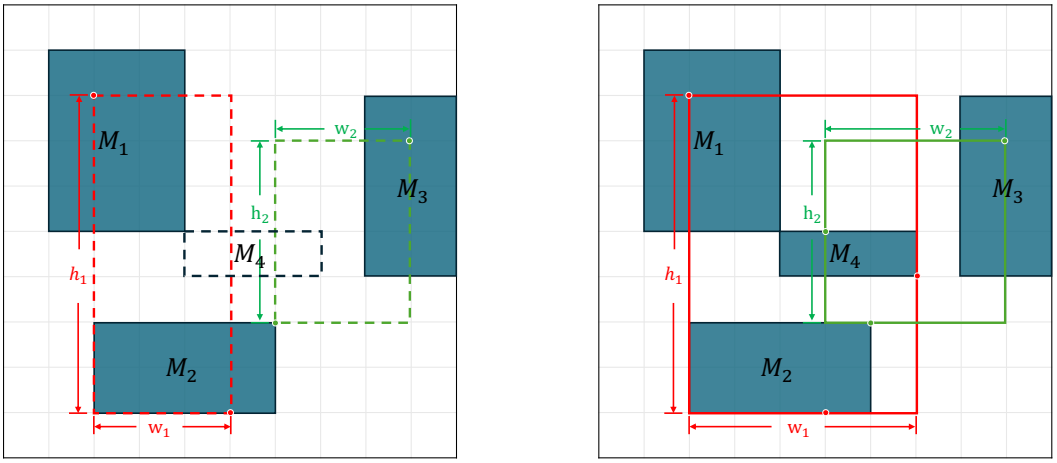
Table A1: Statistics of the ISPD2005 benchmark circuits used in our evaluation.

Circuit	Macros	Macros placed	Hard macros	Standard cells	Nets	Pins	Area util. (%)
adaptec1	543	543	63	210904	221142	944063	55.62
adaptec2	566	566	159	254457	266009	1069482	74.46
adaptec3	723	723	201	450927	466758	1875039	61.51
adaptec4	1329	1329	92	494716	515951	1912420	48.62
bigblue1	560	560	32	277604	284479	1144691	31.58
bigblue2	23084	256	52	534782	577235	2122282	32.43
bigblue3	1293	1293	138	1095519	1123170	3833218	66.81
bigblue4	8170	1024	52	2169183	2229886	8900078	35.68

Although these circuits originate from the ISPD2005 challenge, **they should not be interpreted as sharing a single repeated layout pattern.** Their physical and netlist formations vary sharply. The total macro count ranges from 543 to 23084, the standard-cell count ranges from roughly 2.1×10^5 to 2.2×10^6, the pin count spans almost an order of magnitude, and utilization ranges from 31.58% to 74.46%. These differences change the placement search landscape, including available whitespace, macro blockage structure, fixed-terminal pressure, net bounding-box behavior, and the interaction between macro positions and later standard-cell placement. Therefore, strong results across all eight circuits do not merely indicate that a method fits one benchmark source. They provide evidence that the method remains effective across heterogeneous circuit formations, supporting the robustness and generalizability of DiversePlace under substantially different placement instances.

B Illustration of HPWL Update

In this section, we illustrate how the addition of a macro updates the bounding boxes used for HPWL calculation in Figure B1. Specifically, when a macro is added to a partial placement, each affected net recomputes its bounding box from the extrema of its placed pins, with the HPWL increment equal to the resulting change in half-perimeter.



(a) Before placing M_4. **(b)** After placing M_4.

Figure B1: **Adding M_4 changes affected nets' HPWL bounding boxes.** In partial placement (a), dashed outlines mark M_4's candidate location (black) and the current net bounding boxes (red and green). In (b), M_4 enlarges the affected pin sets; new extrema expand or shift the recomputed boxes, changing each net's half-perimeter contribution. The two affected nets have total HPWL $h_1 + w_1 + h_2 + w_2$, with (h_i, w_i) denoting the i-th net bounding box's height and width.

C Additional Experiment Results and Figures

Figure C1 reports the full cumulative-best HPWL trajectories behind the aggregate results in Section 5. The curves expose both final placement quality and convergence behavior across the complete training budget. Table C1 complements Table 1 with best-observed HPWL at 1K, 2K, and 10K episodes and top-k archive statistics across all eight benchmarks. Its 10K archive HPWL and full-placement diversity values match Table 2.

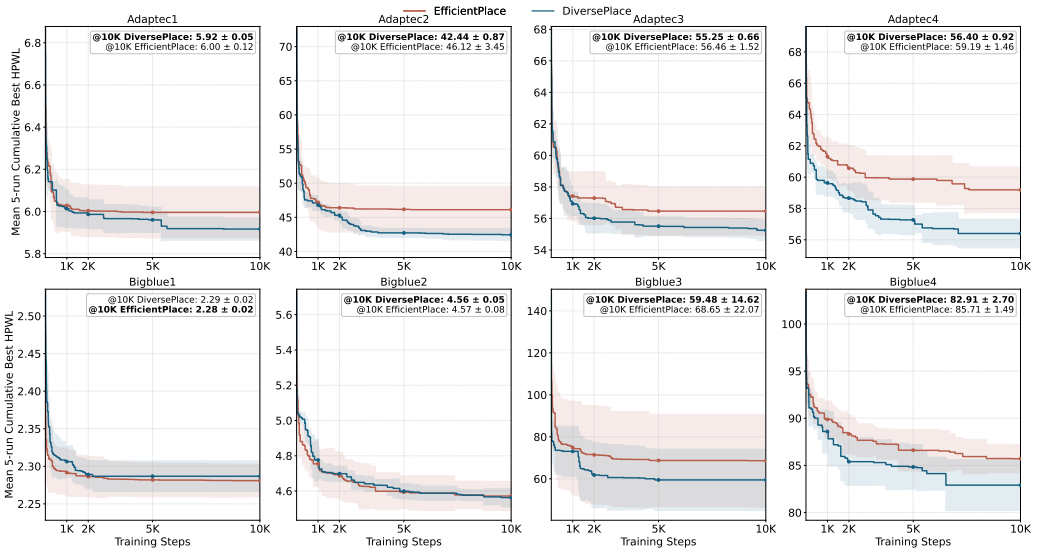


Figure C1: Cumulative-best HPWL trajectories of DiversePlace and EfficientPlace reconstructed from the original run logs. Lower is better.

Table C1: Summary of best-observed HPWL and top-k archive statistics. Each dataset has two rows: DiversePlace (D) and EfficientPlace (E). B denotes best-observed HPWL and follows Table 1. T5/T10/T20 denote mean HPWL within the top-5/top-10/top-20 archives; D5/D10/D20 denote full-placement diversity at 10K. The 10K archive HPWL and diversity values match Table 2. Values are rounded to two decimals. Lower HPWL and higher diversity win; winners after rounding are shown in **bold**.

Dataset	Method	1K HPWL				2K HPWL				10K HPWL				10K Diversity		
		B	T5	T10	T20	B	T5	T10	T20	B	T5	T10	T20	D5	D10	D20
Adaptec1	D	**6.01**	6.08	6.11	6.16	**5.99**	6.04	6.06	6.09	**5.91**	**5.96**	**5.98**	**6.00**	**3.98**	**3.89**	**3.86**
	E	6.03	**6.05**	**6.05**	**6.07**	6.00	**6.02**	**6.02**	**6.03**	6.00	6.00	6.00	6.01	2.13	2.17	2.17
Adaptec2	D	**46.81**	47.92	48.60	49.40	**44.59**	**45.18**	**45.50**	**45.97**	**42.44**	**42.77**	**42.92**	**43.08**	**1.83**	**1.77**	**1.75**
	E	47.19	**47.66**	**47.94**	**48.24**	46.40	46.53	46.66	46.81	46.12	46.24	46.33	46.42	1.48	1.51	1.53
Adaptec3	D	**56.93**	**57.47**	**57.95**	58.62	**55.48**	**56.14**	**56.43**	**56.80**	**55.25**	**55.82**	**56.01**	**56.25**	**4.20**	**4.28**	**4.34**
	E	57.42	57.84	58.11	**58.41**	57.29	57.53	57.73	57.97	56.46	56.87	57.05	57.24	3.41	3.62	3.68
Adaptec4	D	**59.64**	**60.67**	**61.16**	**61.74**	**58.66**	**59.10**	**59.44**	**59.94**	**56.40**	**57.16**	**57.58**	**58.00**	**10.89**	**11.42**	**11.69**
	E	61.32	62.07	62.42	62.85	60.57	61.05	61.35	61.76	59.19	59.88	60.21	60.55	8.56	8.75	8.82
Bigblue1	D	2.31	2.31	2.32	2.32	2.29	2.29	**2.29**	2.30	2.29	2.29	2.29	2.30	**3.80**	**3.82**	**3.84**
	E	**2.29**	**2.30**	**2.31**	**2.31**	2.29	2.29	2.30	2.30	**2.28**	2.29	2.29	**2.29**	2.79	2.76	2.71
Bigblue2	D	4.78	4.86	4.91	4.96	4.70	4.73	4.76	4.80	**4.56**	4.61	4.63	4.66	**2.28**	**2.20**	**2.49**
	E	**4.75**	**4.80**	**4.82**	**4.86**	**4.69**	4.73	**4.75**	**4.77**	4.57	4.61	**4.62**	**4.64**	0.62	0.62	0.67
Bigblue3	D	**73.10**	82.14	86.78	93.63	**61.81**	**64.58**	**65.80**	**67.29**	**59.48**	**60.92**	**61.77**	**62.75**	**4.57**	**4.66**	**4.65**
	E	74.96	**76.57**	**77.87**	**79.22**	71.43	72.70	73.41	74.40	68.65	69.75	70.32	70.98	3.37	3.38	3.47
Bigblue4	D	**88.60**	**90.14**	**91.09**	**92.17**	**85.40**	**87.45**	**88.72**	**90.01**	**82.91**	**85.05**	**86.03**	**86.89**	**6.13**	**6.24**	**6.26**
	E	89.88	91.36	92.21	93.18	88.33	89.41	90.08	90.91	85.71	86.79	87.28	87.90	4.42	4.27	4.28

D Mixed-Size Placement and Congestion Evaluation

RUDY overflow metric. Rectangular uniform wire density (RUDY) is first computed as a two-dimensional map over the final layout. Each bin in this map estimates the local routing utilization induced by net bounding boxes; values above one indicate demand beyond the normalized routing capacity of that bin. To summarize the whole layout with one scalar, we report the normalized

overflow of this map:

$$\mathrm{RUDYOverflow}(P) = \begin{cases} \frac{\sum_{b\in B}\max(U_b-1,0)}{\sum_{b\in B}U_b}, & \sum_{b\in B}U_b > 0, \\ 0, & \sum_{b\in B}U_b = 0, \end{cases} \tag{11}$$

where P is the evaluated placement, B is the routing-bin grid, and $U_b = \max(U_b^h, U_b^v)$ is the larger of the horizontal and vertical capacity-normalized RUDY utilizations in bin b. Thus, the scalar is not the mean RUDY of the image. It is the total amount by which congested bins exceed the capacity threshold, normalized by total estimated routing demand. Algorithm 1 summarizes the computation.

Algorithm 1. RUDY overflow for one completed placement

Require: Completed mixed-size placement P, netlist N, routing-bin grid B, directional capacities c_h, c_v
Ensure: Scalar RUDY overflow

Initialize horizontal and vertical demand maps $D_b^h, D_b^v \leftarrow 0$ for all $b \in B$
for all nets $n \in N$ **do**
  Collect final pin coordinates of n from P
  if n has fewer than two placed pins **then**
    continue
  end if
  Compute the bounding box of n with width W_n and height H_n
  for all bins b overlapped by the bounding box of n **do**
    Compute overlap area $A_{n,b}$ between bin b and the bounding box
    $D_b^h \leftarrow D_b^h + \omega_n A_{n,b}/(H_n + \epsilon)$
    $D_b^v \leftarrow D_b^v + \omega_n A_{n,b}/(W_n + \epsilon)$
  end for
end for
for all bins $b \in B$ **do**
  $U_b^h \leftarrow D_b^h/(\mathrm{area}(b)c_h)$; $U_b^v \leftarrow D_b^v/(\mathrm{area}(b)c_v)$
  $U_b \leftarrow \max(U_b^h, U_b^v)$
end for
if $\sum_{b\in B} U_b > 0$ **then**
  return $\sum_{b\in B}\max(U_b - 1, 0)/\sum_{b\in B} U_b$
else
  return 0
end if

The RUDY signal is only available after mixed-size placement. During macro-placement training, the standard cells do not yet have final physical locations, so the complete post-refinement pin coordinates required by the RUDY map are unavailable. We therefore use RUDY only as a downstream evaluation and archive-selection signal after DREAMPlace finishes mixed-size placement; it is not used as a training-time reward. All EfficientPlace and DiversePlace candidates use the same routing-bin resolution for a given circuit during RUDY evaluation. Table D1 reports the average evaluation time per completed candidate in our cached runs.

The combined row sums phase-1 DREAMPlace cell placement, phase-2 DREAMPlace mixed-size placement, and RUDY evaluation; it excludes macro-placement training and archive generation. These costs are for evaluating a single completed layout. If RUDY were incorporated directly into RL training as a reward or selection signal, each episode or candidate requiring such feedback would need to run the downstream placement-and-congestion evaluation before the signal becomes available. This would make training prohibitively slow, so we treat RUDY as a time-consuming downstream metric rather than a training-time signal.

Mixed-size placement. We follow a standard two-stage mixed-size placement practice used by recent RL-based placement research [10]. Given macro positions produced by a macro placer, DREAMPlace first performs standard-cell placement with macros fixed and without legalization. The resulting placement initializes a second DREAMPlace mixed-size placement run with global placement, legalization, and detailed placement enabled. The DREAMPlace row in Table D2 is reported accordingly [28, 10]. For each seed and each method, we first take the top-k macro archive ranked by macro-placement HPWL, run the same DREAMPlace mixed-size flow for every candidate in this archive, and then report the best downstream mixed-size HPWL within that archive.

Circuit	adaptec1	adaptec2	adaptec3	adaptec4	bigblue1	bigblue2	bigblue3	bigblue4
RUDY bins	512^2	1024^2	1024^2	1024^2	512^2	1024^2	2048^2	2048^2
RUDY only (min.)	1.85	6.05	8.00	7.05	2.26	4.76	37.75	100.71
DREAMPlace+RUDY (min.)	2.85	7.49	11.12	10.89	3.64	7.58	45.26	118.75

Table D1: Average downstream evaluation cost per completed mixed-size placement candidate. 'RUDY only' measures congestion evaluation after the final phase-2 placement is already available. 'DREAMPlace+RUDY' sums phase-1 DREAMPlace placement, phase-2 DREAMPlace mixed-size placement, and RUDY evaluation. Both EfficientPlace and DiversePlace use the same routing-bin resolution for each benchmark during RUDY evaluation.

Table D2: Mixed-size placement HPWL comparison on ISPD2005 circuits. Values are final DREAMPlace mixed-size HPWL in $\times 10^7$ units; lower is better. The standalone DREAMPlace results follow prior reported values [10]. For EfficientPlace and DiversePlace, we evaluate the corresponding macro placements from Table 1 through the same downstream DREAMPlace mixed-size flow. Results are averaged over five predetermined independent runs with seeds 0–4 and reported as mean $\pm$ sample standard deviation.

Method	adaptec1	adaptec2	adaptec3	adaptec4	bigblue1	bigblue2	bigblue3	bigblue4
DREAMPlace	11.10±1.31	13.84±1.74	17.03±0.99	24.37±1.13	10.06±0.28	–	36.51±0.56	175.86±2.23
Top-5 EfficientPlace+DREAMPlace	10.42±0.35	11.24±1.49	24.20±2.64	**22.66±0.56**	10.41±0.32	14.79±0.13	50.07±1.61	105.65±4.95
Top-5 DiversePlace+DREAMPlace	**9.53±0.44**	**10.82±0.31**	**22.15±1.68**	23.00±1.56	10.41±0.23	**14.68±0.17**	**41.45±4.50**	**105.30±2.78**
Top-10 EfficientPlace+DREAMPlace	10.19±0.41	11.16±1.50	23.41±2.23	22.63±0.57	**10.31±0.37**	14.77±0.12	49.83±1.80	105.38±5.16
Top-10 DiversePlace+DREAMPlace	**9.48±0.38**	**10.65±0.48**	**21.94±1.47**	**22.61±1.70**	10.36±0.18	**14.66±0.14**	**39.82±3.60**	**105.12±2.83**

Table D2 shows that the benefits of a diverse macro-placement archive carry over to mixed-size placement. Across both top-5 and top-10 archives, DiversePlace yields better DREAMPlace-refined HPWL than EfficientPlace on most circuits, indicating a greater chance of finding a candidate that better survives standard-cell placement and refinement. Downstream-aware selection from the archive can recover solutions that would be missed by a single proxy-optimal placement. Thus, quality-guided diversity supports exploration and provides a practical interface between macro optimization and subsequent physical-design stages. Figure D1 visualizes mixed-size placements for every dataset.

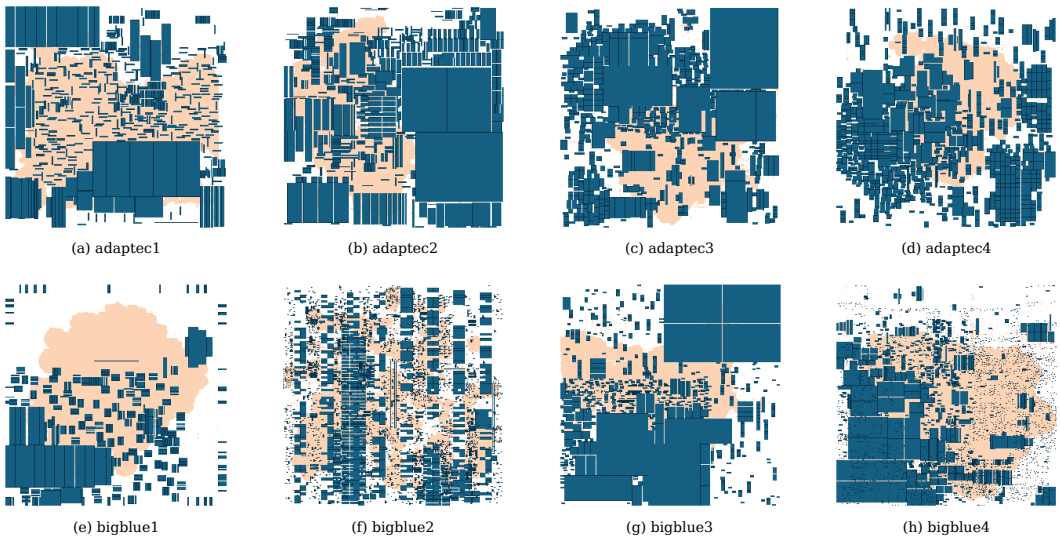


Figure D1: Visualization of mixed-size placement results for the ISPD2005 benchmarks in our evaluations. Macros are shown in blue, and standard cells are shown in orange.

Congestion evaluation. We further investigate the downstream congestion results using RUDY overflow [29]. The detailed evaluation settings are reported above, and EfficientPlace and DiversePlace use the same routing-bin resolution for each benchmark. Table D3 reports the normalized congestion comparison, where EfficientPlace is normalized to 1.00 on each dataset and lower values

are better. Although our primary optimization objective is HPWL, DiversePlace achieves lower normalized RUDY overflow across the evaluated circuits. These results indicate that the quality-guided diverse archive can have a positive effect on downstream routability, suggesting that the additional layout diversity is useful beyond improving archive variation alone.

Table D3: Normalized RUDY congestion comparison. Lower values are better. The MaskPlace row reports the normalized reference values from [10] for reference only.

Method	adaptec1	adaptec2	adaptec3	adaptec4	bigblue1	bigblue2	bigblue3	bigblue4
MaskPlace reference	1.17	2.09	1.18	1.20	4.43	–	1.70	–
Top-5 EfficientPlace	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Top-5 DiversePlace	**0.57**	**0.94**	**0.75**	**0.90**	**0.70**	**0.73**	**0.74**	**0.80**

E Sensitivity Analysis of Key Parameters

The gated quality threshold and QD-map resolution are the two key hyperparameters that directly control our quality-guided diversity mechanism. The gated quality threshold determines which candidate placements are considered competitive enough to enter the diversity-seeking archive, while the QD-map resolution controls how finely the archive separates different displacement modes. We conduct a focused sensitivity study on *adaptec1*, using HPWL as the placement-quality metric and D_{20}^{full} and D_{20}^{PPO} as the archive-diversity metrics, same as in Section 5. All sensitivity rows use cached top-20 placement archives; the "DiversePlace default" row uses our main-experiment setting, with gated quality epsilon 0.05 and 8 QD-map bins, reporting rounded values from Tables 1 and 2.

Table E1: Gated quality ϵ sensitivity on *adaptec1*.

ϵ	HPWL	D_{20}^{full}	D_{20}^{PPO}
0.03	5.95±0.06	**3.99**	1.70
0.05 default	**5.91**±0.05	3.86	1.81
0.07	5.93±0.07	3.98	**1.85**
0.10	5.97±0.06	3.84	1.67

Table E2: QD-map bin sensitivity on *adaptec1*.

QD bins	HPWL	D_{20}^{full}	D_{20}^{PPO}
6	**5.88**±0.02	3.60	1.52
8 default	5.91±0.05	3.86	**1.81**
10	5.94±0.05	**3.93**	1.70
12	5.98±0.09	3.86	1.70

Table E1 shows that relaxing the gate does not monotonically improve the archive. A tighter $\epsilon = 0.03$ gate yields the highest full-placement diversity, while $\epsilon = 0.07$ gives the highest PPO-region diversity. The default $\epsilon = 0.05$ achieves the best mean HPWL with competitive diversity. Further relaxation to $\epsilon = 0.10$ degrades HPWL, suggesting that admitting lower-quality candidates to the diversity phase can weaken placement optimization. We use $\epsilon = 0.05$ across datasets as a conservative default that preserves the quality filter and encourages nontrivial diversity within the archive.

Table E2 shows the expected tradeoff between map granularity and optimization pressure. Six bins give the best HPWL on *adaptec1* but reduce both full-placement and PPO-region diversity relative to the default. Ten or twelve bins encourage broader archive coverage, although stronger diversity pressure can degrade HPWL. The default eight bins balance strong HPWL with the highest PPO-region diversity among the tested settings. We therefore fix eight bins across datasets in all main experiments rather than select a circuit-specific value. The six-bin result suggests additional headroom from per-instance tuning, but we report the uniform default to maintain a fair and reusable protocol for evaluation across datasets.

F Mixed-Size Visualizations for Diversity Comparison

Figure F1 extends the macro-only archive visualization in Figure 5 to the corresponding mixed-size placements. We use exactly the same *adaptec4* macro candidates as Figure 5; after fixing those macro locations, DREAMPlace places the standard cells and refines the complete mixed-size layout. Figure F2 applies the same visualization protocol to *adaptec3*, providing an additional example of how archive-level macro diversity carries over after mixed-size placement. Together, these views show whether the macro-level archive diversity remains visible after standard-cell insertion.

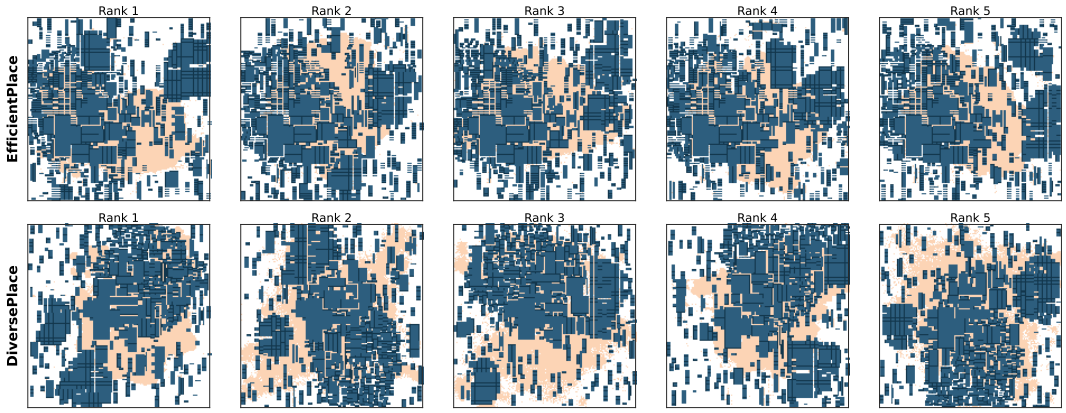


Figure F1: **Mixed-size version of Figure 5 on *adaptec4*.** The macro candidates are exactly the top-5 archive placements shown in Figure 5. Macros are shown in blue and standard cells in orange.

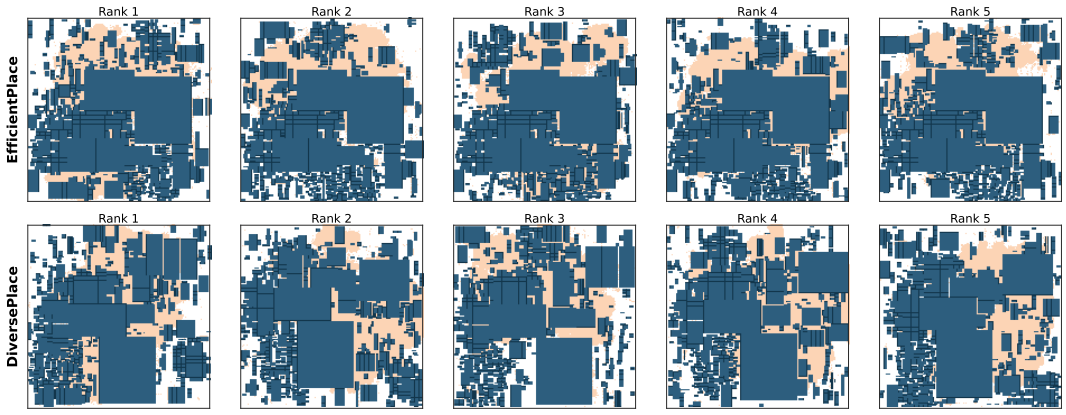


Figure F2: **Mixed-size top-5 archive placements on *adaptec3*.** The top and bottom rows show EfficientPlace and DiversePlace candidates, respectively. Macros are shown in blue, and standard cells are shown in orange.

G Full Placement-and-Routing Evaluation

G.1 HPWL-Only Objective

DiversePlace increases the empirical yield of high-quality downstream candidates by exploring distinct placement modes within a competitive HPWL region. This directly addresses the proxy-driven nature of macro placement: preserving alternative layouts gives downstream implementation more opportunities to find favorable timing–wirelength solutions.

Evaluation protocol. We evaluate DiversePlace and EfficientPlace on *ariane133* using the same official ORFS/OpenROAD flow. Both methods use training seeds 0, 1, and 2, and every seed contributes equally to the reported results. Each run supplies 20 HPWL-ranked candidates, giving both methods the same downstream evaluation budget.

High-quality candidate yield. DiversePlace completes the full flow for 98.33% of its candidates, compared with 86.67% for EfficientPlace. More importantly, its jointly high-quality candidate yield increases from 18.33% to 35.00%. A candidate meets this criterion when its Fmax is at least the median of EfficientPlace's successful candidates and its routed wirelength is at most the corresponding median. These shared thresholds evaluate whether an archive provides candidates that are simultaneously strong in timing and wirelength.

Table G1: Archive utility on *ariane133* across seeds 0, 1, and 2, with 20 candidates per seed. Joint-quality yield includes all evaluated candidates in the denominator. Pareto counts are computed from the pooled successful candidates of both methods.

Archive statistic	EfficientPlace	DiversePlace
Full-flow completion	52/60 (86.67%)	**59/60 (98.33%)**
Jointly high-quality candidates	11/60 (18.33%)	**21/60 (35.00%)**
Candidates on the common Fmax–wirelength Pareto frontier	3	**12**

The nearly doubled joint-quality yield in Table G1 shows that DiversePlace supplies more useful candidates within the same evaluation budget. Its larger contribution to the Pareto frontier further expands the available timing–wirelength choices. Together, these results support the central purpose of the diverse archive: increasing the opportunity to discover a strong downstream implementation through broader coverage of competitive placement modes.

Downstream selection across runs. This higher candidate yield is accompanied by better selected implementations when results are averaged over all three seeds. We compare selecting the lowest-HPWL candidate with selecting the highest-Fmax successful implementation from the Top-20 archive. For each seed, all reported metrics come from the same selected placement.

Table G2: Post-route results on *ariane133*, averaged equally over seeds 0, 1, and 2. All metrics within a seed are measured on the same selected placement.

Selection	Method	Fmax ↑ (MHz)	Setup WNS ↑ (ns)	Setup TNS ↑ (ns)	Routed WL ↓ ($\times 10^6$)
HPWL rank-1	EfficientPlace	182.25	-1.49	-92.67	6.12
HPWL rank-1	DiversePlace	**182.65**	$\mathbf{-1.48}$	$\mathbf{-91.58}$	6.12
Top-20 Fmax	EfficientPlace	183.73	-1.44	-89.75	6.13
Top-20 Fmax	DiversePlace	**184.19**	$\mathbf{-1.43}$	$\mathbf{-88.74}$	**6.11**

Selecting from the DiversePlace archive improves average Fmax, setup slack, and routed wirelength relative to HPWL rank-1 selection. DiversePlace also achieves better mean timing than EfficientPlace under both selection rules and shorter routed wirelength with Top-20 Fmax selection (Table G2). The combination of higher candidate yield and stronger cross-seed selection results demonstrates the practical benefit of preserving multiple competitive layouts for downstream evaluation.

G.2 Composite Quality Objective

We further evaluate whether DiversePlace can provide useful downstream alternatives when placement quality incorporates regularity and logical grouping alongside HPWL. Motivated by regularity-aware refinement in MaskRegulate [30] and expert-guided macro grouping in EXPlace [8], we use the composite objective

$$Q(P) = 0.8\,\widetilde{H}(P) + 0.1\,\widetilde{R}(P) + 0.1\,\widetilde{G}(P), \tag{12}$$

where the three terms measure normalized HPWL, boundary regularity, and hierarchy-based group compactness. Each cost is normalized by its value on the benchmark's fixed legalized reference placement. Both methods use the same composite objective and feasible-action constraints, and candidates are ranked by Q.

Evaluation across seeds and archive sizes. We use the same training seeds, 0, 1, and 2, for both methods and include all three in the reported mean and sample standard deviation. For each seed, we evaluate the composite-ranked Top-5, Top-10, and Top-20 archives through the same ORFS/OpenROAD flow described above. At each archive size, we select the successfully routed candidate with the highest Fmax and report all metrics from that placement.

Table G3: Post-route results from composite-ranked archives on *ariane133*. Entries are mean ± sample standard deviation over seeds 0, 1, and 2. All metrics within a seed and archive size come from the same Fmax-selected placement. Relative improvements are calculated before display rounding.

Archive	Method	Fmax ↑ (MHz)	Setup WNS ↑ (ns)	Setup TNS ↑ (ns)	Routed WL ↓ ($\times 10^6$)
Top-5	EfficientPlace	181.74 ± 5.34	-1.51 ± 0.16	-94.25 ± 9.88	7.54 ± 0.45
Top-5	DiversePlace	$\mathbf{184.05 \pm 0.47}$	$\mathbf{-1.43 \pm 0.01}$	$\mathbf{-89.37 \pm 1.06}$	$\mathbf{6.00 \pm 0.06}$
Top-10	EfficientPlace	182.16 ± 5.83	-1.49 ± 0.18	-93.22 ± 11.06	7.54 ± 0.45
Top-10	DiversePlace	$\mathbf{184.18 \pm 0.42}$	$\mathbf{-1.43 \pm 0.01}$	$\mathbf{-89.21 \pm 1.03}$	$\mathbf{6.02 \pm 0.08}$
Top-20	EfficientPlace	182.20 ± 5.87	-1.49 ± 0.18	-93.37 ± 10.88	7.51 ± 0.42
Top-20	DiversePlace	$\mathbf{184.18 \pm 0.42}$	$\mathbf{-1.43 \pm 0.01}$	$\mathbf{-89.21 \pm 1.03}$	$\mathbf{6.02 \pm 0.08}$

Consistent downstream quality. DiversePlace achieves higher mean Fmax, better setup slack, and shorter routed wirelength at every evaluated archive size (Table G3). Across Top-5, Top-10, and Top-20, it improves mean Fmax by 1.09–1.27% and reduces routed wirelength by 19.83–20.40%. These gains are accompanied by lower variation across runs: its Fmax standard deviation is 0.42–0.47 MHz, compared with 5.34–5.87 MHz for EfficientPlace.

The favorable mean results across archive sizes, together with the smaller across-seed variation, show that DiversePlace provides consistent access to strong downstream implementations under the composite objective. This extends the archive-based formulation to a richer quality definition: quality-guided exploration across distinct placement modes continues to produce useful timing and wirelength choices across repeated runs.

H Broader Impacts

This work studies RL-based macro placement for physical design. Its primary positive impact is to improve the efficiency and quality of chip-design automation. By producing multiple HPWL-competitive macro placements rather than a single proxy-optimal solution, DiversePlace can offer downstream physical-design tools a richer candidate set, potentially reducing manual design iterations and improving the chance of finding layouts with favorable mixed-size placement or congestion behavior. More efficient macro placement may lower computing and engineering costs in early chip design, benefiting academic research and industrial electronic design automation workflows.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Introduction state the scoped contribution as quality-controlled diversity and curriculum PPO for macro placement under an HPWL proxy. The claimed HPWL, runtime, archive-diversity, and downstream-evaluation results are supported by Section 5 and Appendices C–F.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 6 limits DiversePlace to early-stage macro placement with an HPWL-anchored quality gate. The appendix explains why richer downstream signals such as RUDY serve evaluation rather than online training feedback.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper does not present theoretical theorems, lemmas, or proof-based claims. The mathematical expressions define the placement objective, diversity descriptor, QD reward, and evaluation metrics.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Sections 5 and 4.2 specify the benchmark suite, baselines, metrics, seeds, episode budgets, archive construction, and main method components. The appendix provides dataset statistics, full trajectory summaries, downstream DREAMPlace/RUDY protocols, and sensitivity settings for the key diversity parameters.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide the public GitHub repository linked in the paper for access to the code and data used in our evaluation.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper states that PPO is the optimizer, defines the quality-gated diversity reward and expanding-horizon curriculum, and reports the benchmark instances, seeds, budgets, metrics, baselines, and downstream evaluation protocol. Appendix E reports the default quality gate and QD-map resolution and studies their sensitivity.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The main HPWL and mixed-size placement tables report means and sample standard deviations over five predetermined independent seeds, and all seeds are included.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Section 5 reports the CPU/GPU server configuration, GPU memory, one-GPU-per-seed protocol, and training runtimes at the 1K and 2K checkpoints. Appendix D additionally reports per-candidate DREAMPlace and RUDY evaluation costs.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work concerns algorithmic physical-design optimization on public circuit benchmarks. It involves no deployment decisions affecting individuals, surveillance, private personal data, or human subjects.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We include a section in Appendix H to discuss our social impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We release no high-risk pretrained model, image generator, scraped dataset, or dual-use foundation model, and evaluate EDA optimization on public circuit benchmarks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper credits the ISPD2005 benchmark suite, DREAMPlace, Efficient-Place, and other compared methods through proper citations.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: No new dataset, benchmark, pretrained model, or separately released asset is introduced. The paper and appendix document the algorithm and evaluation artifacts.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This work includes no experiments with human subjects and no crowdsourcing, user studies, or annotation labor.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The work does not involve human subjects, so IRB approval or equivalent human-subject review is not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method is a PPO-based macro-placement algorithm with a QD archive and curriculum schedule. It does not use LLMs as a methodological component.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.