DIFFERENTIABLE QUALITY DIVERSITY WITH CROSS-TASK COLLABORATIVE DISCOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

Quality-Diversity (QD) optimization aims to discover collections of high-performing solutions with diverse behaviors. In many applications, however, multiple related QD tasks may need to be solved due to different execution environments. Existing methods typically optimize such multiple tasks independently, without explicitly coordinating their search processes, which can lead to repeated exploration of related regions and redundant rediscovery of useful solutions. In this paper, we propose *Multitask Differentiable Quality-Diversity with Cross-Task Collaborative Discovery* (MTQD-CoD), which jointly optimizes multiple related QD tasks through cross-task collaboration. Specifically, a *subspace-projected repulsion* mechanism encourages task-specific searchers to explore less-overlapping regions, while an *inter-task transfer probability adaptation* mechanism adjusts cross-task transfer probabilities according to the observed archive contributions of transferred candidates, thereby regulating how strongly discoveries are shared across tasks. By reducing redundant exploration while reusing transferable discoveries, MTQD-CoD creates more opportunities to explore undercovered behavioral regions and identify additional high-quality solutions. Experiments on arm repertoire, latent-space illumination, and policy optimization problems demonstrate the effectiveness of MTQD-CoD in discovering diverse and high-quality solutions across multiple related QD tasks.

1 INTRODUCTION

Diversity is increasingly considered in artificial intelligence problems where a single optimized solution may not perform well across operating conditions (Lakshminarayanan et al., 2017; Lehman & Stanley, 2011; Eysenbach et al., 2018; Nam et al., 2021). Related ideas have been explored in robustness learning (Dong et al., 2025), where diverse adversarial examples can expose different model failure modes, and in policy optimization (Eysenbach et al., 2018), where behaviorally distinct policies can support adaptation across different environments. *Quality-Diversity* (QD) optimization (Mouret & Clune, 2015; Pugh et al., 2016; Fontaine et al., 2020) provides a general framework for this setting by discovering solutions that are both high-performing and diverse with respect to predefined behavioral descriptors. Rather than returning a single optimum, QD methods aim to cover different regions of the behavior space with high-quality solutions, forming an archive of diverse solutions.

A key aspect of QD optimization is to continually discover behaviorally distinct solutions while maintaining high solution quality (Mouret & Clune, 2015; Pugh et al., 2016; Fontaine et al., 2020; Pierrot et al., 2022). To do so, QD methods rely on task-specific objectives and behavior descriptors to guide the search toward high-performing solutions across different regions of the behavior space. In practice, however, a single system may need to undergo such QD optimization under multiple related environments. Although the resulting QD tasks can differ in their evaluation criteria and desired behaviors, some useful solution information may remain relevant across tasks. When these tasks are optimized independently, discoveries made during one task's search cannot directly benefit the others, and separate search processes may therefore spend evaluations rediscovering related useful solutions. Such redundant discovery may limit the search from expanding toward unseen behaviorally novel regions, thereby reducing the opportunity to discover additional high-quality and behavioral novel solutions.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

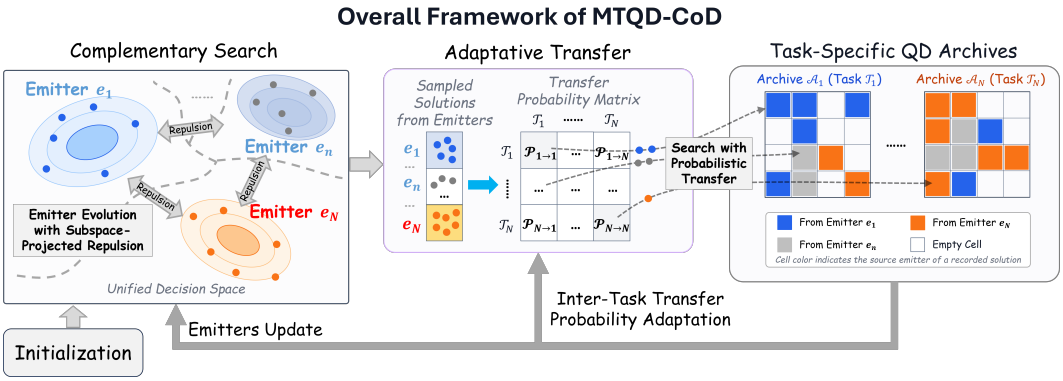

Figure 1: **Overall framework of MTQD-CoD.** Task-specific *emitters* (i.e., the search units of the QD optimizer) jointly search in a unified decision space, where complementary search driven by subspace-projected repulsion to reduce redundant exploration. Discovered solutions are probabilistically transferred across tasks according to adaptively updated transfer probabilities and evaluated for insertion into task-specific QD archives. Archive feedback is then used to update both the emitters and transfer probabilities, forming an iterative cross-task collaborative discovery process.

In this paper, we investigate cross-task collaboration among multiple related QD tasks, with the goal of reducing redundant exploration while preserving task-specific objectives, behavior descriptors, and elite archives. The key idea is to (1) encourage task-specific searches to explore complementary regions of a unified decision space, referred to as *complementary search*, and (2) selectively reuse useful discoveries across tasks through *adaptive transfer*. Building on this principle, we propose *Multitask Differentiable Quality-Diversity with Cross-Task Collaborative Discovery* (MTQD-CoD), a differentiable emitter-based framework that jointly optimizes multiple related QD tasks (Fig. 1). MTQD-CoD realizes *cross-task collaborative discovery* through two coupled mechanisms corresponding to the complementary search and the adaptive transfer. First, we introduce *subspace-projected repulsion* into emitter evolution, encouraging different task-specific emitters to pursue distinct but potentially archive-improving search directions rather than repeatedly exploring similar regions. Second, we develop *inter-task transfer probability adaptation*, which maintains and updates a transfer probability matrix online according to the archive contributions of transferred candidates. This allows discoveries from one task to be probabilistically reused by other tasks, increasing the probability of beneficial transfer channels while suppressing ineffective ones. Together, complementary search broadens the regions collectively explored across tasks, while adaptive transfer allows useful discoveries to benefit multiple task-specific archives, thereby providing more opportunities to discover behaviorally diverse and high-quality solutions. Our key contributions are listed as follows:

- We propose *MTQD-CoD*, a differentiable QD framework that enables multiple related QD tasks to collaboratively discover high quality and novel solutions.
- We develop two strategies for cross-task collaborative discovery: *subspace projected repulsion*, which encourages task-specific searches to pursue complementary regions, and *inter-task transfer probability adaptation*, which regulates the transfer probability of discovered candidate solutions through an adaptively adjusted probability matrix.
- We evaluate MTQD-CoD on three categories of QD problems, including arm repertoire, latent space illumination, and policy optimization. The results show that cross-task collaborative discovery can produce competitive or improved archives of diverse and high-performing solutions.

2 Preliminaries

Multitask Quality Diversity Optimization. We first formulate the multitask QD optimization problem (Hou et al., 2025). Consider a set of N related QD tasks, denoted by $\{\mathcal{T}_n\}_{n=1}^{N}$. Each task $\mathcal{T}_n$ is defined by a task-specific objective function $f_n : \mathcal{X} \to \mathbb{R}$ and a task-specific behavior descriptor $\mathbf{b}_n : \mathcal{X} \to \mathcal{M}_n$, where $\mathcal{X} \subseteq \mathbb{R}^d$ denotes a unified decision space and $\mathcal{M}_n \subseteq \mathbb{R}^K$ denotes the behavior space of task $\mathcal{T}_n$. Specifically, $\mathcal{M}_n$ is partitioned into M cells, i.e., $\mathcal{M}_n = \bigcup_{m=1}^{M} \mathcal{M}_{n,m}$.

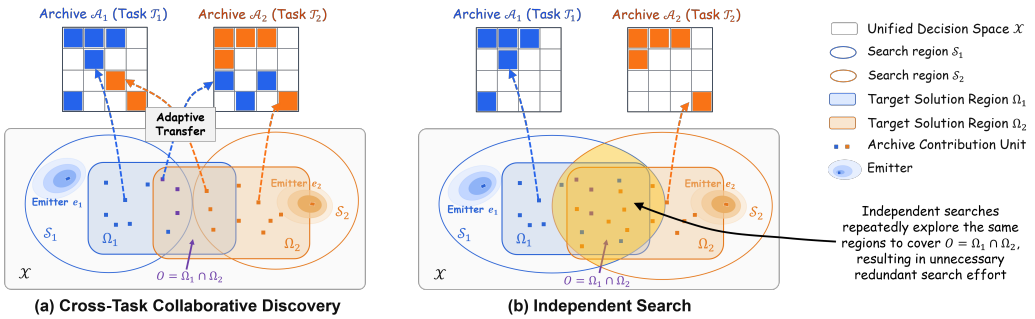

Figure 2: **Cross-task collaborative discovery under overlapping target solution regions.** (a) Cross-task collaborative discovery promotes complementary exploration across task-specific search regions and selectively transfer potential archive-improving discoveries. (b) Independent QD search may repeatedly explore the overlapping region, resulting in redundant search effort.

For each task $\mathcal{T}_n$, QD aims to maintain an archive containing one elite solution for each behavior cell. The elite associated with cell $\mathcal{M}_{n,m}$ is defined as

$$\mathbf{x}^*_{n,m} = \arg\max_{\mathbf{x}\in\mathcal{X}} f_n(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{b}_n(\mathbf{x}) \in \mathcal{M}_{n,m}. \tag{1}$$

The optimal archive for task $\mathcal{T}_n$ is then given by $\mathcal{A}_n = \{\mathbf{x}^*_{n,m}\}_{m=1}^{M}$, where empty cells may remain unfilled if no feasible solution has been discovered. Multitask QD seeks to construct the task-specific optimal archives $\{\mathcal{A}_n\}_{n=1}^{N}$ jointly, leveraging the potential relationship among multiple tasks.

Differentiable Quality Diversity Algorithms. Differentiable QD algorithms (Fontaine & Nikolaidis, 2021; Zhao et al., 2025; Batra et al., 2024) assume that the objective function f and behavior descriptor $\mathbf{b} = (b_1, \ldots, b_K)$ are differentiable. They exploit the gradients ∇f and ∇b_k to guide the search of an *emitter* (i.e., the search unit in differentiable QD algorithms) (Tjanaka et al., 2023). A representative method is CMA-MEGA, whose emitter maintains a search point $\mathbf{x}$ and samples a population of gradient coefficient vectors $\mathbf{c}^{(i)} = (c_0^{(i)}, c_1^{(i)}, \ldots, c_K^{(i)})$ from a maintained Gaussian distribution. Each candidate is generated by taking a gradient step in the decision space: $\mathbf{x}^{(i)} = \mathbf{x} + c_0^{(i)}\nabla f(\mathbf{x}) + \sum_{k=1}^{K} c_k^{(i)}\nabla b_k(\mathbf{x})$. After evaluation, candidates are ranked by their archive improvement, such as whether they discover an empty cell or improve the objective value of an existing elite. This ranking is used in two coupled updates. First, a gradient-free optimizer, typically CMA-ES (Hansen et al., 2001) or xNES (Wierstra et al., 2014; Glasmachers et al., 2010), updates the Gaussian distribution over coefficient vectors so that future samples are more likely to produce archive-improving gradient combinations. Second, the search point $\mathbf{x}$ is updated along a weighted combination of the successful gradient steps, optionally using optimizers such as SGD (Robbins & Monro, 1951) or Adam (Kingma & Ba, 2014). In this way, differentiable QD combines first-order local information with gradient-free adaptation over gradient coefficients, enabling the algorithm to improve quality while exploring the behavior space.

3 Heuristic Explanation of Cross-Task Collaborative Discovery

In this section, we proposed a heuristic explanation for the proposed cross-task collaborative discovery mechanism. In general, cross-task collaborative discovery consists of two coupled components: *complementary search* and *adaptive transfer*. Complementary search encourages task-specific searchers to explore less-overlapping regions, while adaptive transfer controls the transfer of candidates across tasks through dynamically adjusted transfer probabilities. To clarify why the cross-task collaborative discovery work, we consider a simplified two-task setting with $n \in \{1, 2\}$. For each task, we define two abstract regions in the unified decision space. The *target solution region* Ω_n is the smallest connected region that covers the optimal archive $\mathcal{A}_n$, and the *search region* $\mathcal{S}_n$ denotes the region that can be reached by the searcher of task $\mathcal{T}_n$. Under the above abstractions, we discuss two representative cases: *overlapping* target solution regions and *disjoint* target solution regions.

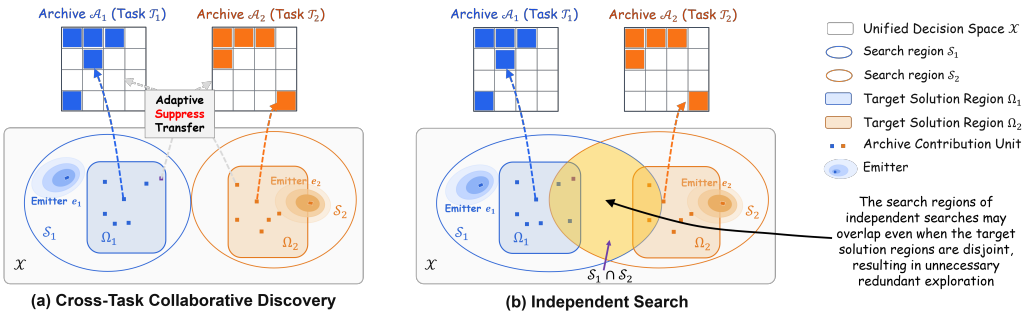

Figure 3: **Cross-task collaborative discovery under disjoint target solution regions.** (a) Cross-task collaborative discovery encourages complementary exploration across task-specific search regions and adaptively suppresses ineffective cross-task transfer, allowing each task to focus more on its own target region. (b) Under independent QD search, the search regions may overlap even when the target solution regions are disjoint, resulting in unnecessary redundant exploration.

Cross-Task Collaborative Discovery under Overlapping Target Solution Regions. Suppose the target solution regions Ω_1 and Ω_2 partially overlap, as illustrated in Fig. 2. Under independent QD search, each task allocates search effort to covering its own target region, yielding search regions $\mathcal{S}_n$ with $\Omega_n \subseteq \mathcal{S}_n$ for $n \in \{1, 2\}$. When $\Omega_1 \cap \Omega_2 \neq \emptyset$, the two tasks may independently explore the shared region, resulting in potentially redundant search effort, as shown in Fig. 2(b). In contrast, collaborative discovery encourages more complementary exploration, biasing each task toward regions that are less extensively explored by the other task, as illustrated in Fig. 2(a). Adaptive transfer further enables solutions discovered by one task to be evaluated on the other, allowing useful discoveries from shared regions to be reused across task-specific archives without requiring independent rediscovery. Together, complementary exploration and cross-task reuse may improve the effective coverage of the joint search process and provide additional opportunities to discover diverse, high-quality solutions across related QD tasks.

Cross-Task Collaborative Discovery under Disjoint Target Solution Regions. We next consider the case where Ω_1 and Ω_2 have little or no overlap, as illustrated in Fig. 3. In this setting, each task primarily benefits from exploring its own target region, and discoveries from the other task are less likely to improve its archive. Nevertheless, even when the target regions are disjoint, independent searches without explicit searching bias may still explore overlapping regions, resulting in $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \emptyset$, as illustrated in Fig. 3(b). In contrast, as shown in Fig. 3(a), cross-task collaborative discovery adapts to this setting through the same two mechanisms. Complementary search encourages the two tasks to explore different regions, reducing unnecessary redundant search, while adaptive transfer reduces cross-task transfer when transferred candidates provide limited archive improvement. Together, these mechanisms allocate more search effort to task-specific unexplored regions and reduce ineffective redundant evaluations, thereby improving the potential efficiency of discovering archive-improving solutions across the two tasks.

We further provide an idealized analysis of cross-task collaborative discovery in QD search using a coupon-collector abstraction. The analysis offers an intuitive view of how collaborative discovery may reduce redundant search effort, with details provided in **Appendix A**.

4 Multitask Differentiable Quality Diversity with Cross-Task Collaborative Discovery

4.1 Overall Framework of MTQD-CoD

Motivated by cross-task collaborative discovery, we propose *MTQD-CoD*, a multitask differentiable QD algorithm that jointly optimizes related tasks in a unified decision space. MTQD-CoD adopts an *emitter*-based paradigm (Tjanaka et al., 2023), with each emitter serving as a task-specific searcher, and coordinates them through *subspace-projected repulsion* for complementary search and *inter-task transfer probability adaptation* for archive-contribution-driven adaptive transfer. The overall

framework is illustrated in Fig. 1, with the implementation details and pseudocode provided **Appendix D**.

- **Initialization.** Given N related QD tasks, MTQD-CoD initializes one archive $\mathcal{A}_n$ and one differentiable emitter e_n for each task $\mathcal{T}_n$. It also initializes an inter-task transfer probability matrix $\mathcal{P} \in [0,1]^{N\times N}$, where the n-th row specifies the probability of assigning candidates generated by emitter e_n to different target tasks.
- **Search with Probabilistic Transfer.** At each iteration, emitter e_n queries its search point $\mathbf{x}_n$ and computes the task-specific objective, behavior descriptor, and their gradients. It then generates candidate solutions along differentiable QD search directions: $\mathbf{x}_n^{(i)} = \mathbf{x}_n + c_{n,0}^{(i)} \nabla f_n(\mathbf{x}_n) + \sum_{k=1}^{K} c_{n,k}^{(i)} \nabla b_{n,k}(\mathbf{x}_n)$, where $\mathbf{c}_n^{(i)} = (c_{n,0}^{(i)}, \ldots, c_{n,K}^{(i)})$ denotes the i-th gradient coefficients sampled via the Gaussian distribution maintained by e_n. Each candidate is assigned a target task $\tau_n^{(i)}$ sampled from the categorical distribution defined by the n-th row of $\mathcal{P}$. The candidate is then evaluated on task $\mathcal{T}_{\tau_n^{(i)}}$, and the resulting objective value and behavior descriptor are used to update the corresponding archive $\mathcal{A}_{\tau_n^{(i)}}$.
- **Emitter Evolution with Subspace Projected Repulsion.** After archive updates, MTQD-CoD evolves the emitters based on the feedback from the updated archive and the single of the subspace projected repulsion. Particularly, the subspace projected repulsion encourages emitters associated with different tasks to search different regions of the unified decision space. The detail about the emitter evolution process is given in Section 4.2.
- **Inter-Task Transfer Probability Adaptation.** The transfer matrix $\mathcal{P}$ is updated according to archive contributions produced by transferred candidate solutions. Successful transfers lead to higher transfer probabilities, whereas ineffective transfers suppresses the transfer probabilities. The detailed procedure is provided in Section 4.3.

The above process is repeated until the termination criterion is met. The algorithm returns the task-specific archives $\{\mathcal{A}_n\}_{n=1}^{N}$, each of which stores diverse elite solutions evaluated under its own objective function and behavior descriptor.

4.2 Emitter Evolution with Subspace Projected Repulsion

In differentiable QD, an emitter drives the search by generating candidate solutions via sampling a coefficient vector $\mathbf{c}$ that linearly combines the objective and behavior-descriptor gradients. For task $\mathcal{T}_n$, let $\mathbf{G}_n = [\nabla_{\mathbf{x}} f_n(\mathbf{x}_n), \nabla_{\mathbf{x}} b_{n,1}(\mathbf{x}_n), \ldots, \nabla_{\mathbf{x}} b_{n,K}(\mathbf{x}_n)]$ be the local gradient basis at the search point $\mathbf{x}_n$. Candidate solutions are generated in the corresponding local linear subspace as $\tilde{\mathbf{x}} = \mathbf{x}_n + \mathbf{c}_n^{\top}\mathbf{G}_n$, where $\mathbf{c}_n \sim \mathcal{N}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$. After evaluation, the archive feedback is passed to the gradient-free optimizer maintained by the emitter, such as CMA-ES in CMA-MEGA (Fontaine & Nikolaidis, 2021), which updates $(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$ and thereby adapts the local search direction. MTQD-CoD builds on this coefficient-space view and introduces *subspace projected repulsion* to promote complementary search across tasks. Specifically, we impose repulsion in the coefficient space of each emitter. For task $\mathcal{T}_n$, we first project the search distributions of all other tasks into the local coefficient space of emitter e_n, allowing e_n to identify regions already covered by other emitters. We then convert this overlap information into a rank-based repulsion signal and return it to the gradient-free optimizer that updates $(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$. Thus, subspace projected repulsion preserves the standard differentiable QD update interface while steering each emitter toward archive-improving yet less redundant search directions.

For task $\mathcal{T}_n$, we define a local coefficient space induced by its gradient matrix $\mathbf{G}_n$. Specifically, we use the ridge-regularized projection $\mathbf{P}_n = \left(\mathbf{G}_n^{\top}\mathbf{G}_n + \lambda \mathbf{I}\right)^{-1} \mathbf{G}_n^{\top}$, where $\lambda > 0$ is a small regularization coefficient. A candidate $\mathbf{x}_n^{(i)}$ generated by emitter e_n is represented in this local coefficient space as $\bar{\mathbf{c}}_n^{(i)} = \mathbf{P}_n(\mathbf{x}_n^{(i)} - \mathbf{x}_n)$. To measure whether this candidate overlaps with the regions likely to be explored by other tasks, we project their search distributions into the same coefficient space. For task $\mathcal{T}_j$ ($j \neq n$), suppose its emitter samples coefficients from $\mathbf{c}_j \sim \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$. Using the local linear approximation $\mathbf{x} \approx \mathbf{x}_j + \mathbf{c}_j^{\top}\mathbf{G}_j$, the induced distribution of task $\mathcal{T}_j$ in the coefficient space of $\mathcal{T}_n$ is $\bar{\mathbf{c}}_{j\to n} \sim \mathcal{N}(\boldsymbol{\mu}_{j\to n}, \boldsymbol{\Sigma}_{j\to n})$, where

$$\boldsymbol{\mu}_{j\to n} = \mathbf{P}_n(\mathbf{x}_j - \mathbf{x}_n) + \mathbf{P}_n\mathbf{G}_j\boldsymbol{\mu}_j, \qquad \boldsymbol{\Sigma}_{j\to n} = \mathbf{P}_n\mathbf{G}_j\boldsymbol{\Sigma}_j(\mathbf{P}_n\mathbf{G}_j)^{\top} + \varepsilon\mathbf{I}.$$

Here, $\varepsilon > 0$ stabilizes the covariance. This projection allows $\mathbf{c}_n^{(i)}$ to be compared with the regions likely to be explored by other task-specific emitters.

After projection, each candidate generated by emitter e_n is scored according to its Mahalanobis distance from the projected search distributions of other tasks:

$$s_{\text{rep}}(\mathbf{c}_n^{(i)}) = \sum_{j \neq n} \left(\bar{\mathbf{c}}_n^{(i)} - \boldsymbol{\mu}_{j\to n}\right)^\top \boldsymbol{\Sigma}_{j\to n}^{-1} \left(\bar{\mathbf{c}}_n^{(i)} - \boldsymbol{\mu}_{j\to n}\right).$$

A larger s_{rep} indicates that the candidate solution is farther from regions currently explored by other tasks and is therefore preferred by the complementary search mechanism. To incorporate this signal without changing the underlying emitter optimizer, we combine it with the original QD archive improvement feedback by rank aggregation:

$$v_{\text{comb}}^{(i)} = (1-\alpha)\,\text{rank}_{\text{MFEA}}\left(v_{\text{qd}}^{(i)}\right) + \alpha\,\text{rank}\left(s_{\text{rep}}(\mathbf{c}_n^{(i)})\right),$$

where $v_{\text{qd}}^{(i)}$ is the archive feedback of candidate $\mathbf{x}_n^{(i)}$, $\alpha \in [0,1]$ controls the strength of repulsion, and $\text{rank}(\cdot)$/$\text{rank}_{\text{MFEA}}(\cdot)$ are rank operators. Both rank operators are computed within the candidate batch, with better ranks assigned to better candidate solutions on the archive improvement. It should be noted that, probabilistic transfer may evaluate candidates generated by emitter e_n on other target tasks, so their archive feedback can come from different tasks. Thus, we use the MFEA-style ranking operator (Gupta et al., 2015) $\text{rank}_{\text{MFEA}}(\cdot)$ to provide a unified comparison across such cross-task evaluated candidate solutions and feed their target-archive contributions back to the source emitter. The combined feedback is then returned as the fitness to the gradient-free optimizer of emitter e_n to update the Gaussian distribution over coefficient vectors $\mathbf{c}_n$:

$$(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) \leftarrow \text{Update}_{\text{GFO}}\left(\{\mathbf{c}_n^{(i)}, v_{\text{comb}}^{(i)}\}_{i=1}^{B}\right).$$

Then, the resulting coefficient vector directions are used to guide the update of the emitter search point. As a result, repulsion-aware ranking guides the evolution of the emitter indirectly through $\mathbf{c}_n$, favoring archive-improving directions that reduce overlap with other task-specific searches.

4.3 Inter-Task Transfer Probability Adaptation

Subspace projected repulsion promotes complementary exploration across tasks, but it does not specify the strength of inter-task transfer. MTQD-CoD therefore introduces an bandit-like adaptive mechanism to update the inter-task transfer probability matrix $\mathcal{P}$ online. Here, $\mathcal{P}_{ij}$ denotes the probability that a candidate generated by source task $\mathcal{T}_i$ is evaluated on target task $\mathcal{T}_j$, with each row defining a categorical distribution: $\sum_{j=1}^{N} \mathcal{P}_{ij} = 1$. We parameterize $\mathcal{P}$ by logits $\boldsymbol{\Theta} \in \mathbb{R}^{N\times N}$ using a row-wise softmax: $\mathcal{P}_{ij}^{(t)} = \left(\exp(\Theta_{ij}^{(t)}/\tau)\right) / \left(\sum_{l=1}^{L} \exp(\Theta_{il}^{(t)}/\tau)\right)$, where $\tau > 0$ is the softmax temperature. At each iteration, candidate solutions generated by task $\mathcal{T}_i$ are assigned target tasks according to the categorical distribution defined by the i-th row of $\mathcal{P}^{(t)}$.

To update $\mathcal{P}$, we measure the utility of each transfer channel by its contribution to the target archive. Let $\mathcal{B}_{ij}^{(t)}$ be the set of candidates generated by $\mathcal{T}_i$ and evaluated on $\mathcal{T}_j$ at iteration t. The archive-contribution reward is defined as

$$C_{ij}^{(t)} = \frac{1}{\max(|\mathcal{B}_{ij}^{(t)}|, 1)} \sum_{\mathbf{x}\in\mathcal{B}_{ij}^{(t)}} \mathbb{I}\{\mathbf{x} \text{ improves } \mathcal{A}_j\}.$$

This reward estimates how often candidates from source task $\mathcal{T}_i$ improve the archive of target task $\mathcal{T}_j$. Since transfer is not always beneficial, we bias the update toward task independence when cross-task contribution is weak, i.e., $R_{ij}^{(t)} = C_{ij}^{(t)} - \beta\left(\mathcal{P}_{ij}^{(t)} - \delta_{ij}\right)$, where $\beta \geq 0$ controls the strength of the independence bias and δ_{ij} is the Kronecker delta. We then maintain an exponential moving average of the reward, i.e., $G_{ij}^{(t)} = \rho G_{ij}^{(t-1)} + R_{ij}^{(t)}$, where $\rho \in [0,1]$ is the decay factor and $G_{ij}^{(0)} = 0$. For each source task, a row-wise baseline is subtracted to obtain the relative advantage of selecting target task $\mathcal{T}_j$, i.e., $A_{ij}^{(t)} = G_{ij}^{(t)} - \frac{1}{N}\sum_{k=1}^{N} G_{ik}^{(t)}$. Finally, the logits are updated according to the

transfer advantage: $\Theta_{ij}^{(t+1)} = \Theta_{ij}^{(t)} + \eta A_{ij}^{(t)} \left(1 - \mathcal{P}_{ij}^{(t)}\right)$, where $\eta > 0$ is the learning rate. Following this update, if candidate solutions from $\mathcal{T}_i$ consistently improve $\mathcal{A}_j$, the advantage $A_{ij}^{(t)}$ becomes positive and $\mathcal{P}_{ij}$ increases. Otherwise, weak archive contribution together with the independence bias suppresses unnecessary transfer.

5 EXPERIMENTS

In this section, we assess MTQD-CoD on arm repertoire, latent-space illumination, and policy optimization problems. Our experiments aim to answer two questions: (i) Can the proposed cross-task collaborative discovery improve the capability of differentiable QD algorithms on searching novel, diverse, and high quality solutions? and (ii) Does it generalize robustly across different types of QD problems?

Benchmarks. We evaluate MTQD-CoD on three groups of multitask QD benchmarks. The first group comprises three two-task arm repertoire problems (Arm-1 to Arm-3) (Fontaine & Nikolaidis, 2021; Zhao et al., 2025); the second consists of three two-task latent space illumination problems (LSI-1 to LSI-3); and the third includes three two-task policy optimization problems based on Walker2D, Ant, and Humanoid MuJoCo environments, following the setting of (Batra et al., 2024). Detailed benchmark specifications are provided in **Appendix C**.

Baselines. For the first two problem groups, we compare MTQD-CoD with four differentiable QD baselines: CMA-MEGA (Fontaine & Nikolaidis, 2021), CMA-MAEGA (Zhao et al., 2025), OG-MAP-Elites (Fontaine & Nikolaidis, 2021), and OMG-MEGA (Fontaine & Nikolaidis, 2021). Since CMA-MAEGA augments CMA-MEGA with an *annealed archive* to improve exploration and archive coverage, and this archive mechanism is orthogonal to our collaborative discovery strategy, we evaluate two MTQD-CoD variants: *MTQD-CoD (ME)*, using the standard CMA-MEGA archive, and *MTQD-CoD (MAE)*, using the annealed archive introduced in CMA-MAEGA. For the policy optimization problems, where QD search is commonly integrated with reinforcement learning, we instantiate MTQD-CoD on top of PPGA (Batra et al., 2024) by incorporating collaborative discovery into its policy search procedure. We compare this variant against its single-task backbone PPGA, as well as two additional QD-based policy optimization methods, DCG-ME (Faldor et al., 2023; 2025) and QD-PG (Pierrot et al., 2022). Implementation details are provided in **Appendix D**.

Performance Metrics. We primarily evaluate performance using behavior-space coverage and the normalized QD score (Pugh et al., 2016). Coverage is the percentage of occupied archive cells and measures the extent of behavioral diversity discovered by an algorithm, whereas the normalized QD score summarizes the quality of the discovered archive solutions. Detailed definitions and calculations are provided in **Appendix D**.

5.1 RESULTS ON ARM REPERTOIRE PROBLEMS

Table 1 summarizes the results on the three two-task arm repertoire problems. Overall, the CMA-based methods substantially outperform OG-MAP-Elites and OMG-MEGA, with MTQD-CoD achieving the strongest results among the compared methods. In particular, *MTQD-CoD (MAE)* obtains the highest coverage on all six tasks and the highest normalized QD score on five tasks, while tying with CMA-MAEGA on Arm-2/Task 2. Compared with their corresponding single-task backbones, both MTQD-CoD variants show consistent improvements: *MTQD-CoD (ME)* outperforms CMA-MEGA on all six tasks in both coverage and normalized QD score, while *MTQD-CoD (MAE)* improves upon CMA-MAEGA in coverage on all six tasks and in normalized QD score on five tasks. These results suggest that cross-task collaborative discovery can complement the underlying differentiable QD search across both standard and annealed archive settings. The improvements are most evident on Arm-1, where both MTQD-CoD variants consistently improve over their respective backbones across the two tasks and both evaluation metrics. This observation is consistent with the larger overlap between the target solution regions of the two Arm-1 tasks: complementary search can reduce repeated exploration of shared regions, while adaptive transfer provides more opportunities for discoveries from one task to improve the archive of the other. The gains become smaller on Arm-2 and Arm-3, where the target-region overlap is reduced, although MTQD-CoD generally retains an advantage over its corresponding backbone. We further observe that the relative gains of *MTQD-CoD (ME)* over CMA-MEGA are generally larger than those of *MTQD-CoD*

Table 1: Results of OG-MAP-Elites, OMG-MEGA, CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE) on Arm Repertoire Problems.

Problem / Task	OG-MAP-Elites	OMG-MEGA	CMA-MEGA	CMA-MAEGA	MTQD-CoD (ME)	MTQD-CoD (MAE)
Coverage ± Std. Dev. (20 Independent Runs)						
Arm-1 / Task 1	$60.60\% \pm 0.34\%$	$60.68\% \pm 0.31\%$	$73.18\% \pm 0.81\%$	$75.57\% \pm 0.54\%$	$74.54\% \pm 0.72\%$	$\mathbf{75.96\% \pm 0.49\%}$
Arm-1 / Task 2	$60.66\% \pm 0.32\%$	$60.53\% \pm 0.26\%$	$72.87\% \pm 0.68\%$	$75.79\% \pm 0.70\%$	$74.48\% \pm 0.77\%$	$\mathbf{75.91\% \pm 0.60\%}$
Arm-2 / Task 1	$33.39\% \pm 0.12\%$	$33.38\% \pm 0.21\%$	$37.10\% \pm 0.66\%$	$38.62\% \pm 0.36\%$	$37.59\% \pm 0.91\%$	$\mathbf{38.92\% \pm 0.67\%}$
Arm-2 / Task 2	$33.35\% \pm 0.19\%$	$33.42\% \pm 0.19\%$	$37.03\% \pm 0.63\%$	$38.71\% \pm 0.42\%$	$37.47\% \pm 1.10\%$	$\mathbf{38.72\% \pm 0.58\%}$
Arm-3 / Task 1	$53.06\% \pm 0.18\%$	$52.96\% \pm 0.22\%$	$59.21\% \pm 0.61\%$	$61.46\% \pm 0.57\%$	$60.40\% \pm 0.90\%$	$\mathbf{61.79\% \pm 0.91\%}$
Arm-3 / Task 2	$53.02\% \pm 0.24\%$	$52.86\% \pm 0.33\%$	$59.54\% \pm 0.55\%$	$61.49\% \pm 0.76\%$	$60.21\% \pm 0.54\%$	$\mathbf{61.67\% \pm 0.71\%}$
Normalized QD Score ± Std. Dev. (20 Independent Runs)						
Arm-1 / Task 1	3.02 ± 0.02	3.03 ± 0.02	3.66 ± 0.04	3.78 ± 0.03	3.73 ± 0.04	$\mathbf{3.80 \pm 0.02}$
Arm-1 / Task 2	3.03 ± 0.02	3.02 ± 0.01	3.64 ± 0.03	3.79 ± 0.03	3.72 ± 0.04	$\mathbf{3.80 \pm 0.03}$
Arm-2 / Task 1	1.67 ± 0.01	1.67 ± 0.01	1.86 ± 0.03	1.93 ± 0.02	1.88 ± 0.05	$\mathbf{1.95 \pm 0.03}$
Arm-2 / Task 2	1.66 ± 0.01	1.67 ± 0.01	1.85 ± 0.03	1.94 ± 0.02	1.87 ± 0.05	$\mathbf{1.94 \pm 0.03}$
Arm-3 / Task 1	2.65 ± 0.01	2.64 ± 0.01	2.96 ± 0.03	3.07 ± 0.03	3.02 ± 0.04	$\mathbf{3.09 \pm 0.05}$
Arm-3 / Task 2	2.65 ± 0.01	2.64 ± 0.02	2.98 ± 0.03	3.07 ± 0.04	3.01 ± 0.03	$\mathbf{3.08 \pm 0.04}$

(MAE) over CMA-MAEGA. A plausible explanation is that the annealed archive already provides CMA-MAEGA with stronger exploration and archive coverage, leaving less room for additional improvement. The smaller marginal gains under the annealed archive therefore appear consistent with combining cross-task collaborative discovery with an already stronger QD optimizer.

5.2 Results on Latent Space Illumination Problems

Table 2: Results of OG-MAP-Elites, OMG-MEGA, CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE) on Latent Space Illumination Problems.

Problem / Task	OG-MAP-Elites	OMG-MEGA	CMA-MEGA	CMA-MAEGA	MTQD-CoD (ME)	MTQD-CoD (MAE)
Normalized QD Score ± Std. Dev. (10 Independent Runs)						
LSI-1 / Task 1	-3.7923 ± 0.5817	-2.8722 ± 0.5757	22.8217 ± 2.2282	33.2011 ± 1.0429	25.5443 ± 1.3311	$\mathbf{34.1672 \pm 1.9471}$
LSI-1 / Task 2	-4.5455 ± 0.3517	-4.8617 ± 0.5090	41.0092 ± 3.2573	52.0936 ± 1.8816	42.9505 ± 2.9097	$\mathbf{52.7537 \pm 1.9933}$
LSI-2 / Task 1	-3.2490 ± 0.9058	-2.5677 ± 0.7819	22.9688 ± 2.1214	33.0985 ± 1.2655	25.6951 ± 2.7559	$\mathbf{33.9424 \pm 1.4013}$
LSI-2 / Task 2	-3.7308 ± 0.8624	-3.8739 ± 1.2187	20.6306 ± 1.8659	30.1189 ± 1.7364	23.9331 ± 2.2806	$\mathbf{31.6171 \pm 1.1712}$
LSI-3 / Task 1	-3.5109 ± 0.7382	-3.4290 ± 0.8200	23.3168 ± 1.8057	33.1225 ± 1.7871	23.7158 ± 4.0033	$\mathbf{34.5995 \pm 1.2375}$
LSI-3 / Task 2	-5.4530 ± 0.7431	-5.9928 ± 0.9834	41.1783 ± 2.5209	50.5114 ± 1.2712	43.4250 ± 3.1422	$\mathbf{50.7580 \pm 2.0915}$

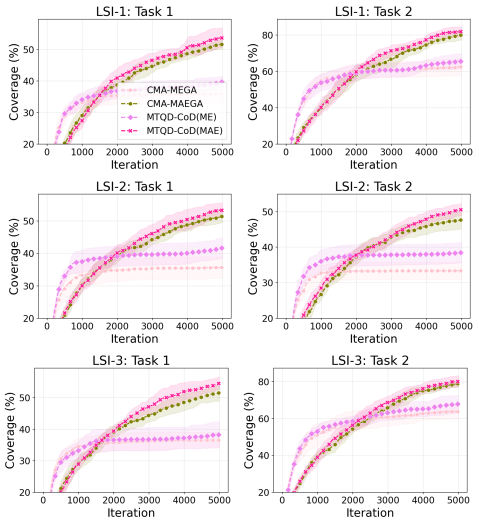

Figure 4: Comparison of coverage convergence trends averaged over 10 independent runs on latent space illumination problems.

Table 2 reports the normalized QD scores on the latent space illumination problems. *MTQD-CoD (MAE)* achieves the highest score on all six tasks, while *MTQD-CoD (ME)* consistently outperforms CMA-MEGA, with the clearest gains on both LSI-2 tasks. The consistency across the two archive variants suggests that the improvement is not specific to a particular archive design, while the smaller gains over CMA-MAEGA are consistent with the stronger exploration already provided by its annealed archive. The coverage convergence curves in Fig. 4 provide similar insights. Cross-task collaborative discovery improves coverage faster on the tasks where its final gains are more pronounced, while remaining broadly comparable to the corresponding backbone on the others. This pattern is consistent with the intended role of cross-task collaborative discovery, which coordinates task-specific searches to reduce redundant exploration and allocate more evaluations to previously under-explored regions, thus improving the discovery capability of the algorithm.

5.3 Results on Policy Optimization Problems

Table 3 reports the normalized QD scores on the policy optimization problems. MTQD-CoD achieves the highest scores on both Walker2D and Humanoid and consistently improves over its

Table 3: Results of QD-PG, DCG-ME, PPGA and MTQD-CoD on Policy Optimization Problems.

Algorithms	Walker2D / Task 1	Walker2D / Task 2	Ant / Task 1	Ant / Task 2	Humanoid / Task 1	Humanoid / Task 2
	Normalized QD Score ± Std. Dev. (5 Independent Runs)					
QD-PG	378.09 ± 13.92	397.04 ± 14.63	322.85 ± 6.80	323.46 ± 18.16	178.03 ± 7.87	186.89 ± 8.09
DCG-ME	609.99 ± 55.77	790.70 ± 61.01	$\mathbf{371.73 \pm 2.10}$	$\mathbf{371.60 \pm 10.28}$	554.78 ± 218.68	693.03 ± 32.95
PPGA	872.29 ± 133.50	832.01 ± 198.91	191.96 ± 43.24	215.24 ± 37.32	2122.03 ± 1200.09	2104.80 ± 1295.32
MTQD-CoD	$\mathbf{1030.01 \pm 121.89}$	$\mathbf{1100.58 \pm 152.98}$	257.67 ± 53.35	278.34 ± 78.30	$\mathbf{3157.58 \pm 929.56}$	$\mathbf{3491.02 \pm 1281.75}$

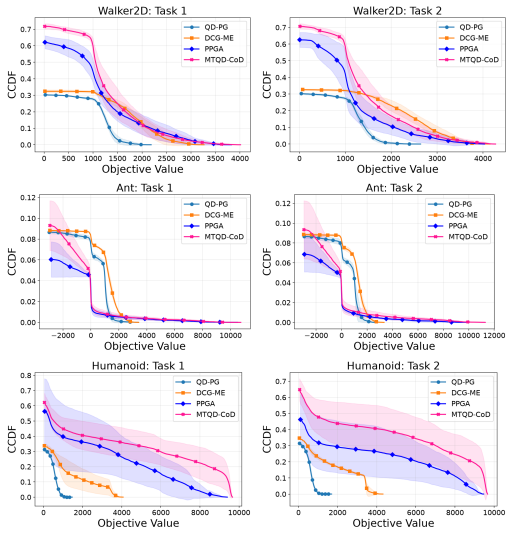

Figure 5: CCDF plots of the objective distributions of archive policies obtained by QD-PG, DCG-ME, PPGA, and MTQD-CoD.

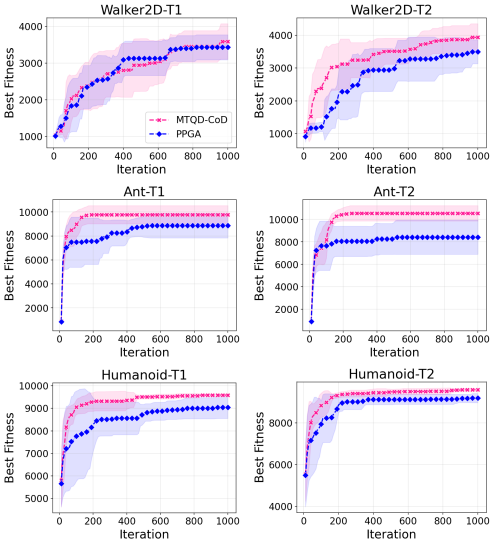

Figure 6: Comparison of best-return convergence trends averaged over 5 independent runs of PPGA and MTQD-CoD.

PPGA backbone on all six tasks, suggesting that cross-task collaborative discovery can provide additional benefits beyond the underlying QD optimizer. We additionally report the Complementary Cumulative Distribution Function (CCDF) curves (Batra et al., 2024; Grillotti et al., 2024) in Fig. 5 to compare solution quality and diversity. On Walker2D, MTQD-CoD shows a broader distribution on the left side of the CCDF, suggesting greater solution diversity. On Ant, despite slightly lower coverage, it exhibits a longer high-return tail, indicating a stronger tendency to discover high-quality policies. On Humanoid, MTQD-CoD improves over the baselines across both diversity and high-return regions. The best-return convergence curves in Fig. 6 further show that MTQD-CoD generally discovers high-performing policies earlier and reaches higher best return values than PPGA. This advantage is also evident on Humanoid, which has a substantially higher-dimensional state space, suggesting that cross-task collaborative discovery remains beneficial in more complex control environments.

Remark: *In addition to the results presented above, we provide further experiments and analyses in* ***Appendix E*** *to* ***Appendix H****.*

6 Conclusion

In this work, we introduced cross-task collaborative discovery for multitask differentiable QD optimization. We developed MTQD-CoD as one concrete realization of cross-task collaborative discovery, jointly optimizing related QD tasks in a unified decision space while preserving task-specific objectives, behavior descriptors, and archives. Experiments on arm repertoire, latent-space illumination, and policy optimization problems show that MTQD-CoD achieves favorable performance in archive coverage and solution quality compared with the corresponding baselines. Ablation and transfer-dynamics analyses reported in the Appendices further indicate that the two components, i.e., complementary search and adaptive transfer, provide distinct benefits under different levels of task relatedness. Overall, the results support cross-task collaborative discovery as an effective approach for improving the discovery of diverse and high-quality solutions in multitask QD optimization.

Reproducibility Statement

We have provided detailed descriptions of the proposed MTQD-CoD framework, algorithmic procedures, benchmark settings, implementation details, and hyperparameter configurations in the main paper and appendices. To further facilitate reproducibility, we will release the source code and experiment configurations upon acceptance of the paper.

AI Use Statement

Generative AI tools were used to assist with language editing and to improve the clarity and readability of the manuscript. They were also used to support the presentation of experimental results, including the plotting, styling, and visual refinement of figures, as well as the formatting and refinement of LaTeX tables. Generative AI was not used to generate, alter, or manipulate experimental data, conduct experiments, or determine the reported results. All AI-assisted text, figures, and tables were reviewed and verified by the authors, who take full responsibility for the final content of the paper.

References

Kavitesh Kumar Bali, Yew-Soon Ong, Abhishek Gupta, and Puay Siew Tan. Multifactorial evolutionary algorithm with online transfer parameter estimation: Mfea-ii. *IEEE Transactions on Evolutionary Computation*, 24(1):69–83, 2019.

Kavitesh Kumar Bali, Abhishek Gupta, Yew-Soon Ong, and Puay Siew Tan. Cognizant multitasking in multiobjective multifactorial evolution: Mo-mfea-ii. *IEEE transactions on cybernetics*, 51(4): 1784–1796, 2020.

Sumeet Batra, Bryon Tjanaka, Matthew Fontaine, Aleksei Petrenko, Stefanos Nikolaidis, and Gaurav Sukhatme. Proximal policy gradient arborescence for quality diversity reinforcement learning. In *International Conference on Learning Representations*, volume 2024, pp. 35884–35904, 2024.

Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. Robots that can adapt like animals. *Nature*, 521(7553):503–507, 2015.

Junhao Dong, Jiao Liu, Xinghua Qu, and Yew-Soon Ong. Confound from all sides, distill with resilience: Multi-objective adversarial paths to zero-shot robustness. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 624–634, 2025.

Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. *arXiv preprint arXiv:1802.06070*, 2018.

Maxence Faldor, Félix Chalumeau, Manon Flageat, and Antoine Cully. Map-elites with descriptor-conditioned gradients and archive distillation into a single policy. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 138–146, 2023.

Maxence Faldor, Félix Chalumeau, Manon Flageat, and Antoine Cully. Synergizing quality-diversity with descriptor-conditioned reinforcement learning. *ACM Transactions on Evolutionary Learning*, 5(1):1–35, 2025.

Liang Feng, Yuxiao Huang, Lei Zhou, Jinghui Zhong, Abhishek Gupta, Ke Tang, and Kay Chen Tan. Explicit evolutionary multitasking for combinatorial optimization: A case study on capacitated vehicle routing problem. *IEEE transactions on cybernetics*, 51(6):3143–3156, 2020.

Matthew Fontaine and Stefanos Nikolaidis. Differentiable quality diversity. *Advances in Neural Information Processing Systems*, 34:10040–10052, 2021.

Matthew C Fontaine, Julian Togelius, Stefanos Nikolaidis, and Amy K Hoover. Covariance matrix adaptation for the rapid illumination of behavior space. In *Proceedings of the 2020 genetic and evolutionary computation conference*, pp. 94–102, 2020.

C Daniel Freeman, Erik Frey, Anton Raichuk, Sertan Girgin, Igor Mordatch, and Olivier Bachem. Brax–a differentiable physics engine for large scale rigid body simulation. *arXiv preprint arXiv:2106.13281*, 2021.

Adam Gaier, Alexander Asteroth, and Jean-Baptiste Mouret. Data-efficient exploration, optimization, and modeling of diverse designs through surrogate-assisted illumination. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 99–106, 2017.

Adam Gaier, Alexander Asteroth, and Jean-Baptiste Mouret. Discovering representations for black-box optimization. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pp. 103–111, 2020.

Theodoros Galanos, Antonios Liapis, Georgios N Yannakakis, and Reinhard Koenig. Arch-elites: Quality-diversity for urban design. In *Proceedings of the genetic and evolutionary computation conference companion*, pp. 313–314, 2021.

Tobias Glasmachers, Tom Schaul, Yi Sun, Daan Wierstra, and Jürgen Schmidhuber. Exponential natural evolution strategies. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, pp. 393–400, 2010.

Luca Grillotti, Maxence Faldor, Borja G León, and Antoine Cully. Quality-diversity actor-critic: learning high-performing and diverse behaviors via value and successor features critics. *arXiv preprint arXiv:2403.09930*, 2024.

Abhishek Gupta, Yew-Soon Ong, and Liang Feng. Multifactorial evolution: Toward evolutionary multitasking. *IEEE Transactions on Evolutionary Computation*, 20(3):343–357, 2015.

Abhishek Gupta, Yew-Soon Ong, Liang Feng, and Kay Chen Tan. Multiobjective multifactorial optimization in evolutionary multitasking. *IEEE transactions on cybernetics*, 47(7):1652–1665, 2016.

Abhishek Gupta, Yew-Soon Ong, and Liang Feng. Insights on transfer optimization: Because experience is the best teacher. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(1):51–64, 2017.

Nikolaus Hansen, Andreas Ostermeier, et al. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.

Saeed Hedayatian and Stefanos Nikolaidis. Soft quality-diversity optimization. In *International Conference on Learning Representations*, volume 2026, pp. 75165–75197, 2026.

Yaqing Hou, Zhaoping Yu, David M. Bossens, Zheng Guo, Ying Wu, Hongwei Ge, and Yew-Soon Ong. Evolutionary heterogeneous multitasking for quality diversity optimization. *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2025. doi: 10.1109/TEVC.2025.3624153.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 4396–4405. IEEE, 2019.

Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33:12104–12114, 2020a.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8107–8116. IEEE, 2020b.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Ariq Koh, Melvin Wong, Jiao Liu, Caishun Chen, Thiago Rios, Stefan Menzel, and Yew Soon Ong. Qd-llms: Quality-diversity optimization with llms for generative design exploration. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 1408–1415, 2026.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Joel Lehman and Kenneth O Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation*, 19(2):189–223, 2011.

Yanchi Li and Wenyin Gong. Multiobjective multitask optimization with multiple knowledge types and transfer adaptation. *IEEE Transactions on Evolutionary Computation*, 29(1):205–216, 2025. doi: 10.1109/TEVC.2024.3353319.

Jiao Liu, Abhishek Gupta, Chinchun Ooi, and Yew-Soon Ong. Extremo: Transfer evolutionary multiobjective optimization with proof of faster convergence. *IEEE Transactions on Evolutionary Computation*, 29(1):102–116, 2024.

Jiao Liu, Abhishek Gupta, and Yew-Soon Ong. Bayesian inverse transfer in evolutionary multiobjective optimization. *ACM Transactions on Evolutionary Learning*, 4(4):1–27, 2025a.

Jiao Liu, Weiming Liu, Joel Tay Wei En, Caishun Chen, Puay Siew Tan, and Yew-Soon Ong. Optimal transport-based distributional pairing in transfer multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 2025b.

Jiao Liu, Meng Xu, Hua Yu, Chi Xu, Puay Siew Tan, and Yew-Soon Ong. Accelerating two-stage stochastic programming via parametric multitask optimization for microfactory-enabled last-mile manufacturing. *IEEE Transactions on Evolutionary Computation*, 2026.

Aritz D Martinez, Javier Del Ser, Eneko Osaba, and Francisco Herrera. Adaptive multifactorial evolutionary optimization for multitask reinforcement learning. *IEEE Transactions on Evolutionary Computation*, 26(2):233–247, 2021.

Alan Tan Wei Min, Yew-Soon Ong, Abhishek Gupta, and Chi-Keong Goh. Multiproblem surrogates: Transfer evolutionary multiobjective optimization of computationally expensive problems. *IEEE Transactions on Evolutionary Computation*, 23(1):15–28, 2017.

Jean-Baptiste Mouret and Jeff Clune. Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909*, 2015.

Jean-Baptiste Mouret and Glenn Maguire. Quality diversity for multi-task optimization. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pp. 121–129, 2020.

Giung Nam, Jongmin Yoon, Yoonho Lee, and Juho Lee. Diversity matters when learning from ensembles. *Advances in neural information processing systems*, 34:8367–8377, 2021.

Olle Nilsson and Antoine Cully. Policy gradient assisted map-elites. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 866–875, 2021.

Thomas Pierrot, Valentin Macé, Felix Chalumeau, Arthur Flajolet, Geoffrey Cideron, Karim Beguir, Antoine Cully, Olivier Sigaud, and Nicolas Perrin-Gilbert. Diversity policy gradient for sample efficient quality-diversity optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1075–1083, 2022.

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.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pp. 400–407, 1951.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Shenglian Tan, Yong Wang, Guangyong Sun, Tong Pang, and Ke Tang. A surrogate-assisted evolutionary framework for expensive multitask optimization problems. *IEEE Transactions on Evolutionary Computation*, 2024.

Bryon Tjanaka, Matthew C Fontaine, David H Lee, Yulun Zhang, Nivedit Reddy Balam, Nathaniel Dennler, Sujay S Garlanka, Nikitas Dimitri Klapsis, and Stefanos Nikolaidis. pyribs: A bare-bones python library for quality diversity optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 220–229, 2023.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pp. 5026–5033. IEEE, 2012.

Handing Wang, Yaochu Jin, Cuie Yang, and Licheng Jiao. Transfer stacking from low-to high-fidelity: A surrogate-assisted bi-fidelity evolutionary algorithm. *Applied soft computing*, 92: 106276, 2020.

Tingyang Wei, Jiao Liu, Abhishek Gupta, Chin Chun Ooi, Puay Siew Tan, and Yew-Soon Ong. Amortized multi-objective optimization across tasks with generative solution modeling. *arXiv preprint arXiv:2511.09598*, 2025a.

Tingyang Wei, Jiao Liu, Abhishek Gupta, Puay Siew Tan, and Yew-Soon Ong. (θ, θ)-parametric multi-task optimization: Joint search in solution and infinite task spaces. *IEEE Transactions on Evolutionary Computation*, 2025b.

Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, and Jürgen Schmidhuber. Natural evolution strategies. *The Journal of Machine Learning Research*, 15(1):949–980, 2014.

Haofeng Wu, Jiao Liu, Yaochu Jin, and Yew-Soon Ong. Extremo-d: Diffusion transfer for expensive multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, 2026.

Xiaoming Xue, Kai Zhang, Kay Chen Tan, Liang Feng, Jian Wang, Guodong Chen, Xinggang Zhao, Liming Zhang, and Jun Yao. Affine transformation-enhanced multifactorial optimization for heterogeneous problems. *IEEE Transactions on Cybernetics*, 52(7):6217–6231, 2020.

Shihan Zhao, Bryon Tjanaka, Matthew C Fontaine, and Stefanos Nikolaidis. Covariance matrix adaptation map-annealing: Theory and experiments. *ACM Transactions on Evolutionary Learning*, 5(1):1–53, 2025.

Lei Zhou, Liang Feng, Kay Chen Tan, Jinghui Zhong, Zexuan Zhu, Kai Liu, and Chao Chen. Toward adaptive knowledge transfer in multifactorial evolutionary computation. *IEEE transactions on cybernetics*, 51(5):2563–2576, 2020.

A Idealized Analysis on Cross-Task Collaborative Discovery

Idealized Analysis Setup. We provide an idealized analysis to clarify how cross-task collaborative discovery can reduce redundant search effort under the two representative cases discussed in Section 3. The analysis is not intended to characterize the exact dynamics of MTQD-CoD; instead, it isolates the two sources of potential computational benefit: avoiding repeated exploration across task-specific searches and reusing discoveries across tasks.

At a fixed behavior-space resolution, we treat each archive-improving discovery as one *archive-contribution unit*. Constructing the target archive of a task therefore corresponds to collecting all contribution units required by that task. We further discretize the corresponding decision-space search region into *search units* and assume that each search evaluation samples one search unit uniformly. Let $T(M, r)$ denote the number of search evaluations required to discover all r archive-contribution units from a search region containing M units. Under the coupon-collector abstraction, its expectation is

$$\mathbb{E}[T(M, r)] = MH_r, \qquad H_r = \sum_{k=1}^{r} \frac{1}{k}, \tag{2}$$

where H_r is the r-th harmonic number. This abstraction captures that, for a fixed set of required discoveries, a larger search region increases the expected number of evaluations needed to recover them.

We then normalize the computational cost of independently discovering an archive-contribution unit through iterative QD search to one unit. This cost abstracts candidate generation and evaluation, behavior-descriptor computation, gradient calculation, and the iterative search required before an archive-improving solution is obtained. In contrast, once a solution has been discovered by another task, transferring it only requires target-task evaluation and archive insertion, while avoiding the iterative search process that would otherwise be required to independently rediscover the same archive-contributing solution. We denote this relative cost by c_{tr} and assume

$$0 < c_{\mathrm{tr}} \leq 1. \tag{3}$$

Finally, we consider an ideal collaborative setting in which complementary search perfectly avoids unnecessary overlap and adaptive transfer correctly identifies useful or ineffective transfer. These assumptions provide an optimistic lower-bound characterization of the computational cost attainable by collaborative discovery, rather than a guarantee for the practical algorithm.

Idealized Analysis under Overlapping Target Solution Regions. We first consider the case in which the two tasks share archive-contribution units, corresponding to the overlapping target solution region setting discussed in Section 3.

Proposition 1. *Consider two QD tasks $\mathcal{T}_1$ and $\mathcal{T}_2$ with target solution regions Ω_1 and Ω_2, where*

$$O = \Omega_1 \cap \Omega_2 \neq \emptyset.$$

Let

$$E_1 = \Omega_1 \setminus \Omega_2, \qquad E_2 = \Omega_2 \setminus \Omega_1$$

denote their task-exclusive target regions. At a fixed behavior-space resolution, suppose that E_1, E_2, and O contain a_1, a_2, and $c > 0$ archive-contribution units, respectively. Let the task-specific search regions $\mathcal{S}_1$ and $\mathcal{S}_2$ contain $b_1 + d$ and $b_2 + d$ search units, where $d > 0$ denotes the number of units in the overlapping search region $\mathcal{S}_1 \cap \mathcal{S}_2$. Assume that all archive-contribution units required by each task are contained in its corresponding search region.

Under independent QD search, each task must discover all of its task-exclusive units as well as all c shared units independently. The expected computational cost is therefore

$$\mathbb{E}[T_{\mathrm{ind}}] = (b_1 + d)H_{a_1+c} + (b_2 + d)H_{a_2+c}. \tag{4}$$

Under ideal collaborative discovery, suppose complementary search partitions the overlapping search region into two disjoint subsets of sizes d_1 and d_2, with

$$d_1 + d_2 = d,$$

and similarly divides the c shared archive-contribution units into subsets of sizes c_1 and c_2, with

$$c_1 + c_2 = c.$$

Task $\mathcal{T}_n$ directly discovers its $a_n + c_n$ assigned contribution units within a search region containing $b_n + d_n$ units. The remaining shared discoveries are obtained from the other task through cross-task transfer. The resulting idealized cost expectation is

$$\mathbb{E}[T_{\mathrm{cod}}]_{\mathrm{lb}} = (b_1 + d_1)H_{a_1+c_1} + (b_2 + d_2)H_{a_2+c_2} + c\,c_{\mathrm{tr}}. \tag{5}$$

For $0 < c_{\mathrm{tr}} \leq 1$,

$$\mathbb{E}[T_{\mathrm{cod}}]_{\mathrm{lb}} \leq \mathbb{E}[T_{\mathrm{ind}}]. \tag{6}$$

Proof. Under independent QD search, Task $\mathcal{T}_1$ must directly discover all $a_1 + c$ archive-contribution units from a search region containing $b_1 + d$ units. Under ideal collaborative discovery, it only needs to directly discover $a_1 + c_1$ units from the reduced search region containing $b_1 + d_1$ units, because the remaining $c_2 = c - c_1$ shared units are discovered by $\mathcal{T}_2$ and transferred to $\mathcal{T}_1$.

We first quantify the saving due to avoiding these c_2 independent rediscoveries. If the search-region size were kept unchanged at $b_1 + d$, the reduction in expected search cost would be

$$(b_1 + d)\,(H_{a_1+c} - H_{a_1+c_1}) = (b_1 + d) \sum_{k=a_1+c_1+1}^{a_1+c} \frac{1}{k}.$$

Since all required contribution units are contained in the search region, $b_1 + d \geq a_1 + c$. Hence, for every $k \leq a_1 + c$,

$$\frac{b_1 + d}{k} \geq 1.$$

The above summation contains exactly $c - c_1 = c_2$ terms, and therefore

$$(b_1 + d)\,(H_{a_1+c} - H_{a_1+c_1}) \geq c_2. \tag{7}$$

Ideal complementary search further reduces the search-region size from $b_1 + d$ to $b_1 + d_1$. Since $d_1 \leq d$, this can only decrease the expected cost of directly discovering the remaining $a_1 + c_1$ units. Consequently,

$$(b_1 + d)H_{a_1+c} - (b_1 + d_1)H_{a_1+c_1} \geq c_2. \tag{8}$$

The same argument applies to Task $\mathcal{T}_2$. Under collaborative discovery, $\mathcal{T}_2$ directly discovers only $a_2 + c_2$ units, while the remaining c_1 shared units are supplied by $\mathcal{T}_1$. Thus,

$$(b_2 + d)H_{a_2+c} - (b_2 + d_2)H_{a_2+c_2} \geq c_1. \tag{9}$$

Adding Eqs. equation 8 and equation 9, and using $c_1 + c_2 = c$, gives

$$\mathbb{E}[T_{\text{ind}}] - [(b_1 + d_1)H_{a_1+c_1} + (b_2 + d_2)H_{a_2+c_2}] \geq c. \tag{10}$$

Thus, before accounting for transfer, ideal collaborative discovery saves at least c units of expected direct-search cost.

Each of the c shared archive-contribution units is transferred once to the other task, incurring a total transfer cost of $c\,c_{\text{tr}}$. Therefore, using equation 5,

$$\mathbb{E}[T_{\text{ind}}] - \mathbb{E}[T_{\text{cod}}]_{\text{lb}} \geq c - c\,c_{\text{tr}} = c(1 - c_{\text{tr}}).$$

Because $c > 0$ and $0 < c_{\text{tr}} \leq 1$,

$$c(1 - c_{\text{tr}}) \geq 0. \tag{11}$$

Hence,

$$\mathbb{E}[T_{\text{cod}}]_{\text{lb}} \leq \mathbb{E}[T_{\text{ind}}], \tag{12}$$

which proves the proposition. □

Proposition 1 formalizes the two potential benefits illustrated in the case of overlapping target solution regions. Complementary search reduces the amount of overlapping decision-space exploration, while cross-task transfer avoids independently rediscovering archive-contribution units that are useful to both tasks. The gain becomes larger as the amount of useful shared structure increases or the relative transfer cost decreases.

Idealized Analysis under Disjoint Target Solution Regions. We next consider the case where the target solution regions are disjoint. In this setting, cross-task solutions provide little direct archive contribution, but the task-specific search regions may still overlap. Cross-task collaborative discovery can therefore remain useful by reducing redundant exploration, even when transfer itself provides no benefit.

Proposition 2. *Consider two QD tasks $\mathcal{T}_1$ and $\mathcal{T}_2$ satisfying*

$$\Omega_1 \cap \Omega_2 = \emptyset. \tag{13}$$

Suppose their target regions contain a_1 and a_2 archive-contribution units, respectively. Their independent search regions contain

$$|\mathcal{S}_1| = b_1 + d, \qquad |\mathcal{S}_2| = b_2 + d,$$

search units, where $d \geq 0$ represents the overlapping region $\mathcal{S}_1 \cap \mathcal{S}_2$.

Under independent QD search, the expected computational cost is

$$\mathbb{E}[T_{\text{ind}}] = (b_1 + d)H_{a_1} + (b_2 + d)H_{a_2}. \tag{14}$$

Under ideal collaborative discovery, suppose complementary search partitions the overlapping search region into two disjoint subsets of sizes d_1 and d_2, where

$$d_1 + d_2 = d,$$

while preserving all archive-contribution units required by the corresponding task. Since the target regions are disjoint, assume ideal adaptive transfer suppresses cross-task evaluations that provide no archive contribution. The resulting cost becomes

$$\mathbb{E}[T_{\text{cod}}]_{\text{lb}} = (b_1 + d_1)H_{a_1} + (b_2 + d_2)H_{a_2}. \tag{15}$$

Then,

$$\mathbb{E}[T_{\text{cod}}]_{\text{lb}} \leq \mathbb{E}[T_{\text{ind}}]. \tag{16}$$

Proof. Subtracting equation 15 from equation 14 gives

$$\mathbb{E}[T_{\text{ind}}] - \mathbb{E}[T_{\text{cod}}]_{\text{lb}} = (d - d_1)H_{a_1} + (d - d_2)H_{a_2}. \tag{17}$$

Since $d_1 + d_2 = d$, we have

$$d - d_1 = d_2, \qquad d - d_2 = d_1,$$

and therefore

$$\mathbb{E}[T_{\text{ind}}] - \mathbb{E}[T_{\text{cod}}]_{\text{lb}} = d_2 H_{a_1} + d_1 H_{a_2}. \tag{18}$$

Because $d \geq 0$, for $a_1, a_2 > 0$, we have $H_{a_1}, H_{a_2} \geq 0$, and thus

$$d_2 H_{a_1} + d_1 H_{a_2} \geq 0. \tag{19}$$

Hence,

$$\mathbb{E}[T_{\text{cod}}]_{\text{lb}} \leq \mathbb{E}[T_{\text{ind}}]. \tag{20}$$

□

Proposition 2 shows that, when $\Omega_1 \cap \Omega_2 = \emptyset$, independently optimized tasks may spend evaluations exploring the same decision-space regions. Ideal complementary search reduces this overlap and therefore decreases the expected cost of collecting the task-specific archive contributions. In this case, adaptive transfer mainly serves to suppress low-utility cross-task evaluations rather than to reuse shared discoveries.

The idealized analysis above highlights two representative sources of benefit from cross-task collaborative discovery. When the target regions overlap, collaboration can reduce both redundant exploration and repeated rediscovery of shared archive-contributing solutions. When the target regions are disjoint, the main benefit comes from reducing overlap between possible task-specific search processes while limiting ineffective transfer. Although this analysis relies on simplified assumptions and does not capture the full dynamics of practical MTQD-CoD, it provides an intuitive computational explanation for why cross-task collaborative discovery can improve the search under different task relationships.

B Related Work

B.1 Quality Diversity Optimization

QD optimization aims to discover a collection of high-performing solutions that exhibit diverse behaviors, rather than optimizing toward a single best solution (Mouret & Clune, 2015; Pugh et al., 2016; Fontaine et al., 2020). This formulation has been widely applied to domains such as robotics (Cully et al., 2015; Nilsson & Cully, 2021), reinforcement learning (Faldor et al., 2023; Batra et al., 2024), and design optimization (Gaier et al., 2017; Galanos et al., 2021; Koh et al., 2026), where a repertoire of diverse and high-quality solutions can be more useful than a single optimum. A recent line of work further introduces gradient information into QD, leading to differentiable QD

methods (Fontaine & Nikolaidis, 2021). Instead of relying solely on stochastic variation, differentiable QD (Fontaine & Nikolaidis, 2021) exploits gradients of both the objective function and behavior descriptors with respect to the decision variables to construct search directions that simultaneously improve solution quality and explore the behavior space. Representative methods such as CMA-MEGA (Fontaine & Nikolaidis, 2021), and CMA-MAEGA (Zhao et al., 2025) combine these differentiable search directions with evolutionary emitters or adaptive archive mechanisms, and have demonstrated improved search efficiency on high-dimensional QD problems. Related policy-gradient-based approaches (Batra et al., 2024; Hedayatian & Nikolaidis, 2026) further extend differentiable QD to reinforcement learning, where gradients of policy performance and behavioral measures can be used to discover diverse high-performing skills. Despite these advances, existing differentiable QD methods are predominantly formulated for single-task settings, where each search process maintains a task-specific archive under a fixed objective and behavior descriptor. Although prior work (Mouret & Maguire, 2020) has discussed the connection between QD and multitask optimization, it primarily uses QD to address multiple optimization tasks, rather than jointly solving multiple QD tasks, each with its own objective, behavior descriptor, and archive. In many practical settings, however, multiple QD tasks may be related to each other, creating opportunities for cross-task knowledge sharing and coordinated search. Existing single-task QD methods do not explicitly model such cross-task interactions, and, to the best of our knowledge, limited work has considered differentiable QD methods specifically designed for jointly optimizing multiple related QD tasks.

B.2 Evolutionary Multitasking

Evolutionary multitasking (EMT) aims to solve multiple related optimization tasks simultaneously by exploiting cross-task knowledge transfer during evolutionary search (Gupta et al., 2017; 2015). Existing methods generally adopt either implicit or explicit transfer. Implicit approaches, such as the MFEA family (Gupta et al., 2015; 2016; Bali et al., 2019), encode multiple tasks in a unified representation and promote information exchange through cross-task genetic interaction, while adaptive variants further estimate task relatedness to reduce negative transfer (Bali et al., 2020; Zhou et al., 2020; Martinez et al., 2021; Li & Gong, 2025). Explicit transfer methods instead construct mappings between task domains to directly transfer promising solutions or search distributions across tasks (Feng et al., 2020; Xue et al., 2020; Liu et al., 2025b). Related ideas have also been extended to expensive optimization (Tan et al., 2024; Min et al., 2017; Liu et al., 2024; Wang et al., 2020; Liu et al., 2025a; Wei et al., 2025b;a; Wu et al., 2026; Liu et al., 2026). Despite these advances, existing EMT methods are primarily developed for conventional optimization settings, where each task seeks one or a small set of high-quality optima. This objective differs fundamentally from QD optimization, which aims to discover a diverse collection of high-performing solutions distributed across a behavior space. Consequently, directly applying conventional multitask transfer mechanisms to QD does not explicitly account for the discovery of behaviorally diverse or novel solutions, nor does it exploit the gradient structure available in differentiable QD. More importantly, the underlying cross-task coordination principle is different. In conventional multitask optimization, related tasks are often encouraged to share information or move their search distributions closer because they may possess similar high-quality solution regions. In contrast, cross-task collaborative discovery in our setting deliberately discourages overlapping search across tasks, i.e., task-specific searchers are repelled from regions already explored by others so that they cover complementary parts of the unified decision space. This broader collective exploration creates more opportunities to uncover behaviorally novel solutions, while useful discoveries can still be selectively reused through adaptive transfer. Thus, unlike conventional multitask optimization, which primarily exploits similarity to accelerate convergence toward high-quality regions, our method uses cross-task interaction to coordinate complementary discovery while preserving selective reuse of shared solutions.

C Details of Benchmarks

C.1 Arm Repertoire Problems

We construct three multitask arm repertoire problems, denoted as **Arm-1**, **Arm-2**, and **Arm-3**, based on the classical planar revolute-arm repertoire problem (Gaier et al., 2020; Mouret & Maguire, 2020). Each problem contains two related tasks, referred to as **Task 1** and **Task 2**. For all the six tasks, the solution is defined as the joint angles of a planar arm with D revolute joints, i.e.,

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D) \in \mathbb{R}^D$. For each task $\mathcal{T}_n$, the objective f_n evaluates the quality of the arm configuration, while the behavior descriptor $\mathbf{b}_n(\boldsymbol{\theta}) \in \mathbb{R}^2$ specifies the normalized Cartesian position of the end effector. We use a 100×100 grid archive over the common behavior domain $[-1, 1]^2$ for all tasks. Following the standard arm repertoire setting, higher-quality solutions are those with smaller variance among joint angles, encouraging joint rotations to be distributed more evenly across the arm rather than concentrated on a small subset of joints. The three problems differ in how inter-task variation is introduced. In Arm-1, the two tasks use different link lengths and therefore differ in arm morphology. In Arm-2 and Arm-3, the arm morphology is unchanged, while different obstacle configurations are introduced to restrict the regions that the end effector can reach. These constructions produce task pairs with different forms and degrees of relatedness, allowing us to evaluate cross-task collaborative discovery under diverse multitask settings.

Objective Function. For task $\mathcal{T}_n$, let $\bar{\theta} = \frac{1}{D}\sum_{i=1}^{D} \theta_i$ denote the mean joint angle, we maximize the negative variance of the joint angles:

$$f_n(\boldsymbol{\theta}) = -\operatorname{Var}(\boldsymbol{\theta}) = -\frac{1}{D}\sum_{i=1}^{D} \left(\theta_i - \bar{\theta}\right)^2. \tag{21}$$

For tasks without obstacles, this objective is shared across tasks. Its maximum value is 0, attained when all joint angles are identical, and its gradient with respect to θ_j is $\frac{\partial f_n}{\partial \theta_j} = -\frac{2}{D}(\theta_j - \bar{\theta})$. For Arm-2 and Arm-3, we further introduce a task-specific smooth penalty for configurations whose end effector enters an obstacle region. Let c_n denote the obstacle boundary for task $\mathcal{T}_n$. An upper obstacle blocks behaviors with $b_{n,2}(\boldsymbol{\theta}) > c_n$, whereas a lower obstacle blocks behaviors with $b_{n,2}(\boldsymbol{\theta}) < c_n$. The penalty is defined as $p_n(\boldsymbol{\theta}) = 0$ when the end effector remains outside the obstacle and $p_n(\boldsymbol{\theta}) = P(1 + \tanh(s_n))/2$ otherwise, where $P = 500$, $w = 0.05$, and $s_n = (b_{n,2}(\boldsymbol{\theta}) - c_n)/w$ for an upper obstacle or $s_n = (c_n - b_{n,2}(\boldsymbol{\theta}))/w$ for a lower obstacle. The resulting task objective is

$$f_n(\boldsymbol{\theta}) = -\operatorname{Var}(\boldsymbol{\theta}) - p_n(\boldsymbol{\theta}). \tag{22}$$

Behavior Measures. Let $l_{n,d}$ denote the length of link d, and let $R_n = \sum_{d=1}^{D} l_{n,d}$ denote the total arm length. Given $\boldsymbol{\theta}$, the absolute orientation of link d is $\varphi_d = \sum_{j=1}^{d} \theta_j$. The behavior descriptor is defined as

$$\mathbf{b}_n(\boldsymbol{\theta}) = \begin{bmatrix} b_{n,1}(\boldsymbol{\theta}) \\ b_{n,2}(\boldsymbol{\theta}) \end{bmatrix} = \frac{1}{R_n} \begin{bmatrix} \sum_{i=1}^{D} l_{n,d}\cos(\varphi_i) \\ \sum_{i=1}^{D} l_{n,d}\sin(\varphi_i) \end{bmatrix}. \tag{23}$$

Since $b_{n,1}^2 + b_{n,2}^2 \le 1$, the geometrically reachable behaviors of an unconstrained arm lie within the unit disk. Accordingly, for the square archive over $[-1, 1]^2$, the continuous reachable area occupies at most a fraction of $\pi/4 \approx 78.5\%$ of the archive domain. The exact maximum coverage of the discretized archive may differ slightly due to grid resolution and can be further reduced by the obstacles in Arm-2 and Arm-3. For differentiable QD methods, the behavior descriptor is differentiable with respect to the joint angles. Its Jacobian entries are $\frac{\partial b_{n,1}}{\partial \theta_j} = -\frac{1}{R_n}\sum_{i=j}^{d} l_{n,i}\sin(\varphi_i)$ and $\frac{\partial b_{n,2}}{\partial \theta_j} = \frac{1}{R_n}\sum_{i=j}^{d} l_{n,i}\cos(\varphi_i)$.

Problem Configurations. The detailed configurations of the three arm repertoire problems are summarized in Table 4. All tasks use the same decision-space dimensionality and archive resolution, while differing in arm morphology or obstacle configuration.

Arm-1 contains the most similar pair of tasks. The two tasks differ only in a uniform scaling of all link lengths, while sharing the same objective and normalized behavior domain. Because the end-effector coordinates are normalized by the total arm length R_n, this scaling preserves the overall structure of the behavior space while changing the underlying arm morphology. **Arm-2** introduces two opposing obstacle configurations. Task 1 places an upper obstacle at $c_1 = 0$, blocking behaviors with $b_{1,2} \ge 0$, whereas Task 2 places a lower obstacle at $c_2 = 0$, blocking behaviors with $b_{2,2} \le 0$. The two tasks therefore encourage exploration of largely complementary parts of

Table 4: Configurations of the multitask arm repertoire problems. All tasks use $D = 1000$ decision variables, a 100×100 grid archive, and the normalized behavior domain $[-1, 1]^2$.

Problem / Task	$l_{n,i}$	R_n	Obstacle	Boundary c_n	Blocked behaviors
Arm-1 / Task 1	1.0	1000	None	–	–
Arm-1 / Task 2	1.5	1500	None	–	–
Arm-2 / Task 1	1.0	1000	Upper	0	$b_{n,2} \geq 0$
Arm-2 / Task 2	1.0	1000	Lower	0	$b_{n,2} \leq 0$
Arm-3 / Task 1	1.0	1000	Upper	0.5	$b_{n,2} \geq 0.5$
Arm-3 / Task 2	1.0	1000	Lower	-0.5	$b_{n,2} \leq -0.5$

Table 5: Configurations of the multitask latent space illumination problems. All tasks optimize a 7168-dimensional normalized StyleGAN2 W^+ latent code and use a 100×100 grid archive over the behavior domain $[-0.2, 0.2]^2$.

Problem / Task	prompt_n	Behavior axis 1	Behavior axis 2	Latent Dim.
LSI-1 / Task 1	''an European woman''	Child ↔ Elderly	Long hair ↔ Short hair	7168
LSI-1 / Task 2	''an European woman''	Serious ↔ Laughing	Red hair ↔ Blonde hair	7168
LSI-2 / Task 1	''an European woman''	Child ↔ Elderly	Long hair ↔ Short hair	7168
LSI-2 / Task 2	''an East Asian man''	Child ↔ Elderly	Long hair ↔ Short hair	7168
LSI-3 / Task 1	''an European woman''	Child ↔ Elderly	Long hair ↔ Short hair	7168
LSI-3 / Task 2	''an East Asian man''	Serious ↔ Laughing	Red hair ↔ Blonde hair	7168

the behavior space. **Arm-3** introduces a milder contrast between the two tasks. Task 1 blocks behaviors with $b_{1,2} \geq 0.5$, while Task 2 blocks behaviors with $b_{2,2} \leq -0.5$. Consequently, both tasks can explore the central band $-0.5 \leq b_{n,2} \leq 0.5$, while each task additionally retains access to a different side of the behavior space. Compared with Arm-2, Arm-3 therefore provides greater overlap between the behaviors accessible to the two tasks.

C.2 Latent Space Illumination Problems

Inspired by (Fontaine & Nikolaidis, 2021), we construct three multitask latent space illumination problems, i.e., **LSI-1**, **LSI-2**, and **LSI-3**, using a fixed StyleGAN2 (Karras et al., 2020a;b) generator and CLIP-based semantic evaluation (Radford et al., 2021). Each problem contains two related tasks, referred to as **Task 1** and **Task 2**. All six tasks share the same high-dimensional decision space, where a solution is represented by a normalized latent code $\boldsymbol{q} \in \mathbb{R}^{7168}$ corresponding to the W^+ latent space of a StyleGAN2 model trained on FFHQ (Karras et al., 2019) at 256×256 resolution. Specifically, let $\boldsymbol{w} \in \mathbb{R}^{14\times512}$ denote a W^+ latent code. We use the mean latent vector $\boldsymbol{w}_{\text{avg}}$ stored in the StyleGAN2 mapping network and estimate the element-wise standard deviation $\boldsymbol{w}_{\text{std}}$ from 10,000 random latent samples obtained through the mapping network. The normalized decision variable is defined as $\boldsymbol{q} = (\boldsymbol{w} - \boldsymbol{w}_{\text{avg}}) \oslash \boldsymbol{w}_{\text{std}}$, and is mapped back to the W^+ space by $\boldsymbol{w} = \boldsymbol{q} \odot \boldsymbol{w}_{\text{std}} + \boldsymbol{w}_{\text{avg}}$, where $\oslash$ and $\odot$ denote element-wise division and multiplication, respectively. During optimization and evaluation, the StyleGAN2 generator and CLIP ViT-B/32 encoder are kept fixed, and gradients are propagated only with respect to the latent variables. We use constant generator noise during these stages so that each latent code deterministically maps to a generated image. The three latent space illumination problems differ in how inter-task variation is introduced through the target identities and semantic behavior attributes. By varying whether the two tasks share the same target identity and/or the same semantic behavior dimensions, we obtain task pairs with different forms.

Objective Function. For each task $\mathcal{T}_n$, the objective f_n measures the semantic agreement between the generated image and a task-specific target identity specified by a text prompt. Let $I(\boldsymbol{q})$ denote the image generated from $\boldsymbol{q}$, and let $\boldsymbol{e}_I$ and $\boldsymbol{e}_{t,n}$ denote the L2-normalized CLIP embeddings of the generated image and the *target text* for task $\mathcal{T}_n$, respectively. We measure their semantic discrepancy using the spherical distance

$$\mathcal{L}_{\text{sph}}\left(\boldsymbol{e}_I(\boldsymbol{q}), \boldsymbol{e}_{t,n}\right) = 2\left[\arcsin\left(\frac{\left\|\boldsymbol{e}_I(\boldsymbol{q}) - \boldsymbol{e}_{t,n}\right\|_2}{2}\right)\right]^2. \tag{24}$$

The *target text* for each task follows the template

$$\texttt{``A photo of the face of \{prompt_n\}''}, \tag{25}$$

with the $\{\texttt{prompt_n}\}$ identity determined by the corresponding task. To keep the search within a typical region of the normalized latent space, we additionally apply a radial regularization term

$$R(\boldsymbol{q}) = 0.2\left(\max\left(\|\boldsymbol{q}\|_2, q_{\text{norm}}\right) - q_{\text{norm}}\right)^2. \tag{26}$$

Here, q_{norm} is set to 0.35 times the mean L2 norm of the normalized latent codes obtained from the same $10{,}000$ random mapping-network samples, i.e.,

$$q_{\text{norm}} = 0.35\,\mathbb{E}\left[\left\|\frac{\boldsymbol{w} - \boldsymbol{w}_{\text{avg}}}{\boldsymbol{w}_{\text{std}}}\right\|_2\right], \tag{27}$$

where the expectation is approximated empirically using these samples. The raw task loss is then

$$\mathcal{L}_n(\boldsymbol{q}) = \mathcal{L}_{\text{sph}}\left(\boldsymbol{e}_I(\boldsymbol{q}), \boldsymbol{e}_{t,n}\right) + R(\boldsymbol{q}). \tag{28}$$

Since the QD framework maximizes the objective, we transform this loss into the task-specific objective

$$f_n(\boldsymbol{q}) = 10\left(10 - 5\mathcal{L}_n(\boldsymbol{q})\right), \tag{29}$$

which is used as the objective function.

Behavior Measures. For each task $\mathcal{T}_n$, the behavior descriptor $\mathbf{b}_n(\boldsymbol{q}) \in \mathbb{R}^2$ is defined using two CLIP-based semantic axes. Let $\boldsymbol{e}_I(\boldsymbol{q})$ denote the L2-normalized CLIP embedding of the generated image, and let $\boldsymbol{e}^+_{n,i}$ and $\boldsymbol{e}^-_{n,i}$ denote the L2-normalized CLIP embeddings of the positive and negative text prompts for semantic axis i, respectively. We define the angular distance between two normalized embeddings as

$$d(\boldsymbol{e}_1, \boldsymbol{e}_2) = 2\arcsin\left(\frac{\|\boldsymbol{e}_1 - \boldsymbol{e}_2\|_2}{2}\right). \tag{30}$$

The i-th behavior component is then defined as

$$b_{n,i}(\boldsymbol{q}) = d\big(\boldsymbol{e}_I(\boldsymbol{q}), \boldsymbol{e}^+_{n,i}\big) - d\big(\boldsymbol{e}_I(\boldsymbol{q}), \boldsymbol{e}^-_{n,i}\big), \qquad i \in \{1, 2\}. \tag{31}$$

Thus, each behavior dimension captures the relative semantic proximity of the generated image to two contrasting attribute descriptions.

For the first set of semantic axes, we use:

$\boldsymbol{e}^+_{n,1}$: $\texttt{``A photo of \{prompt_n\} as a small child.''}$,

$\boldsymbol{e}^-_{n,1}$: $\texttt{``A photo of \{prompt_n\} as an elderly person.''}$,

$\boldsymbol{e}^+_{n,2}$: $\texttt{``A photo of \{prompt_n\} with long hair.''}$,

$\boldsymbol{e}^-_{n,2}$: $\texttt{``A photo of \{prompt_n\} with short hair.''}$.

For the second set of semantic axes, we use:

$\boldsymbol{e}^+_{n,1}$: $\texttt{``A photo of \{prompt_n\} with serious expression.''}$,

$\boldsymbol{e}^-_{n,1}$: $\texttt{``A photo of \{prompt_n\} with laughing expression.''}$,

$\boldsymbol{e}^+_{n,2}$: $\texttt{``A photo of \{prompt_n\} with red hair.''}$,

$\boldsymbol{e}^-_{n,2}$: $\texttt{``A photo of \{prompt_n\} with blonde hair.''}$.

Here, $\{\texttt{prompt_n}\}$ is replaced by the target identity associated with task $\mathcal{T}_n$. Both behavior dimensions are discretized over the common range $[-0.2, 0.2]$ using a 100×100 grid archive.

Benchmark configurations. The detailed configurations of the three latent space illumination problems are summarized in Table 5. We use two target identity descriptions, C_A = $\texttt{``an European woman''}$ and $C_B$ = $\texttt{``an East Asian man''}$. The three problems differ in whether inter-task variation is introduced through the target identity, the semantic behavior axes, or both. In **LSI-1**, the two tasks share the same target identity C_A but use different semantic behavior axes. Task 1 uses the first set, i.e., age and hair length, whereas Task 2 uses the second set, i.e., facial expression and hair color. In **LSI-2**, the two tasks share the same semantic behavior axes but use different target identities, with Task 1 targeting C_A and Task 2 targeting C_B. In **LSI-3**, both the target identity and the semantic behavior axes differ across the two tasks. These configurations provide task pairs with different forms of semantic relatedness while keeping the generator, latent space dimensionality, and archive configuration fixed.

Table 6: Configurations of the multitask policy optimization problems. Each problem contains two MuJoCo control tasks with identical robot and policy configurations but different ground friction scales.

Problem / Task	Friction Scale	Obs. Dim.	Action Dim.	Behavior Dim.	Archive
Walker2d / Task 1	1.0	17	6	2	50×50
Walker2d / Task 2	0.7	17	6	2	50×50
Ant / Task 1	1.0	27	8	4	10^4
Ant / Task 2	0.7	27	8	4	10^4
Humanoid / Task 1	1.0	227	17	2	50×50
Humanoid / Task 2	0.7	227	17	2	50×50

C.3 Policy Optimization Problems

We construct three multitask MuJoCo-based policy optimization problems (Todorov et al., 2012), denoted as **Walker2d**, **Ant**, and **Humanoid**, based on continuous-control environments in Brax (Freeman et al., 2021). Each problem contains two related tasks, referred to as **Task 1** and **Task 2**. Within each problem, the two tasks use the same robot morphology, observation and action spaces, policy network architecture, and behavior representation, while differing only in the environment dynamics induced by the ground friction coefficient. For each task $\mathcal{T}_n$, a solution corresponds to the parameters of a stochastic control policy $\pi_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ denotes the serialized policy network parameters and constitutes the decision variable optimized by QD. The objective $f_n(\boldsymbol{\theta})$ evaluates the cumulative control performance of the policy in task $\mathcal{T}_n$, while the behavior descriptor $\mathbf{b}_n(\boldsymbol{\theta})$ characterizes the resulting locomotion skill using foot-contact patterns. In this way, the archive aims to retain policies that exhibit diverse locomotion behaviors while maintaining high task performance.

The three problems differ in robot morphology and the dimensionality of the corresponding behavior space. Walker2d and Humanoid use two-dimensional foot-contact descriptors, whereas Ant uses a four-dimensional descriptor corresponding to its four feet. For each robot, Task 1 uses the default ground friction scale of 1.0, while Task 2 uses a reduced friction scale of 0.7. The friction scaling is applied only to the ground geometry, thereby changing the contact dynamics between the robot and the ground without modifying the policy parameterization or behavior definition. This construction provides pairs of closely related control tasks with a controlled shift in environment dynamics, allowing us to evaluate cross-task collaborative discovery in policy parameter space.

Objective function. For each task $\mathcal{T}_n$, the objective $f_n(\boldsymbol{\theta})$ is defined as the episodic return of policy $\pi_{\boldsymbol{\theta}}$ in the corresponding Brax environment:

$$f_n(\boldsymbol{\theta}) = \mathbb{E}_{\pi_{\boldsymbol{\theta}}, \mathcal{T}_n}\left[\sum_{t=0}^{H-1} \gamma^t r_t\right], \tag{32}$$

where $H = 1000$ is the episode horizon and $\gamma = 0.99$ is the discount factor. The environment reward follows the standard Brax locomotion objective and combines forward locomotion, healthy-state reward, and control cost. In general, the per-step reward can be written as

$$r_t = r_t^{\text{forward}} + r_t^{\text{healthy}} - r_t^{\text{ctrl}}. \tag{33}$$

Thus, higher objective values correspond to policies that achieve better locomotion performance under the task-specific contact dynamics. Policy optimization follows the PPGA backbone, where PPO (Schulman et al., 2017) is used to estimate and improve the objective and behavior gradients with respect to the policy parameters. The policy and value networks are shared in architecture across the two tasks of each problem, while their rollouts and objective evaluations are performed in their corresponding environments.

Behavior Measures. For each task $\mathcal{T}_n$, the behavior descriptor $\mathbf{b}_n(\boldsymbol{\theta}) \in [0, 1]^{K_n}$ is defined from the contact patterns between designated robot links and the ground. Each behavior dimension describes the contact behavior of one foot-related link during policy execution, providing a compact representation of the locomotion strategy independently of the accumulated reward. For Walker2d, the behavior descriptor is two-dimensional and is constructed from the contact patterns of `foot` and `foot_left`. For Ant, four behavior dimensions are used, corresponding to the four feet. For

Algorithm 1 MTQD-CoD

Require: Tasks $\{\mathcal{T}_n\}_{n=1}^N$, archives $\{\mathcal{A}_n\}_{n=1}^N$, batch size B, number of iterations T
1: Initialize the inter-task transfer probability matrix $\mathcal{P} \in [0,1]^{N\times N}$ and differentiable emitters $\{e_n\}_{n=1}^N$
2: **for** iteration $t = 1, \dots, T$ **do**
3: **for** task $n = 1, \dots, N$ **do**
4: Query the search point $\mathbf{x}_n$ from emitter e_n
5: Evaluate $f_n(\mathbf{x}_n)$, $\mathbf{b}_n(\mathbf{x}_n)$, $\nabla_{\mathbf{x}} f_n(\mathbf{x}_n)$, and $\nabla_{\mathbf{x}} \mathbf{b}_n(\mathbf{x}_n)$
6: Record the evaluation and gradient information into e_n
7: Generate candidate solutions $\{\mathbf{x}_n^{(i)}\}_{i=1}^B$ from e_n
8: Sample target skill factors $\{\tau_n^{(i)}\}_{i=1}^B$ from the categorical distribution parameterized by the n-th row of $\mathcal{P}$
9: **end for**
10: Evaluate each candidate $\mathbf{x}_n^{(i)}$ on its assigned target task $\mathcal{T}_{\tau_n^{(i)}}$
11: Update the corresponding archives $\{\mathcal{A}_n\}_{n=1}^N$ with the evaluated candidates
12: Evolve emitters $\{e_n\}_{n=1}^N$ using subspace projected repulsion as introduced in **Section 4.2**
13: Update $\mathcal{P}$ according to the adaptation method introduced in **Section 4.3**
14: **end for**
15: **return** Archives $\{\mathcal{A}_n\}_{n=1}^N$

Humanoid, the descriptor is two-dimensional and is based on the contact patterns of `left_shin` and `right_shin`. Hence,

$$\mathbf{b}_n(\boldsymbol{\theta}) = [b_{n,1}(\boldsymbol{\theta}), \dots, b_{n,K_n}(\boldsymbol{\theta})]^\top, \qquad b_{n,i}(\boldsymbol{\theta}) \in [0,1], \tag{34}$$

where $K_n = 2$ for Walker2d and Humanoid and $K_n = 4$ for Ant. The behavior domain is discretized using a grid archive. We use a 50×50 archive for Walker2d and Humanoid, and a 10^4 archive for Ant, corresponding to 10 bins along each of its four behavior dimensions. The same behavior definition and archive discretization are used for both tasks within each problem, so that differences between their discovered skill repertoires arise from the change in environment dynamics rather than from different behavior representations.

Benchmark Configurations. The detailed configurations of the three multitask policy optimization problems are summarized in Table 6. All task pairs share the same robot morphology, policy architecture, observation and action spaces, and behavior descriptor within each problem, and differ only in the ground friction scale. Within each problem, Task 1 represents the default contact dynamics with a friction scale of 1.0, whereas Task 2 reduces the ground friction to 0.7. This modification changes how the robot interacts with the ground and can consequently alter both the achievable locomotion performance and the distribution of discovered foot-contact behaviors. At the same time, the shared policy parameterization and behavior representation preserve substantial structure between the two tasks, making these problems suitable for evaluating cross-task transfer and collaborative discovery under controlled changes in environment dynamics.

D Implementation Details and Parameter Settings

D.1 Implementation Details

In this subsection, we provide further implementation details of MTQD-CoD. For the standard optimization tasks, including the arm repertoire and latent space illumination problems, the implementation follows **Algorithm 1**. Specifically, MTQD-CoD builds on the CMA-MEGA framework with CMA-based differentiable emitters, while incorporating the proposed subspace-projected repulsion and cross-task transfer mechanisms to enable collaborative search across multiple QD tasks. For policy optimization, MTQD-CoD is instantiated on top of PPGA, as summarized in **Algorithm 2**. This implementation differs from the standard version mainly in the construction and update of the differentiable emitters. Following PPGA, each emitter maintains an xNES (Glasmachers et al., 2010; Batra et al., 2024) distribution over coefficients associated with the objective and behavior

Algorithm 2 MTQD-CoD for Policy Optimization

Require: Tasks $\{\mathcal{T}_n\}_{n=1}^N$, archives $\{\mathcal{A}_n\}_{n=1}^N$, batch size B, number of iterations T
1: Initialize the inter-task transfer probability matrix $\mathcal{P} \in [0,1]^{N\times N}$ and PPGA emitters $\{e_n\}_{n=1}^N$
2: **for** iteration $t = 1, \ldots, T$ **do**
3: **for** task $n = 1, \ldots, N$ **do**
4: Query the search policy $\boldsymbol{\theta}_n$ from PPGA emitter e_n
5: Evaluate $f_n(\boldsymbol{\theta}_n)$, $\mathbf{b}_n(\boldsymbol{\theta}_n)$, and their policy gradients via PPO rollouts
6: Record the evaluation and gradient information into e_n
7: Generate candidate policies $\{\boldsymbol{\theta}_n^{(i)}\}_{i=1}^B$ from e_n according to the PPGA arborescent search
8: Sample target skill factors $\{\tau_n^{(i)}\}_{i=1}^B$ from the categorical distribution parameterized by the n-th row of $\mathcal{P}$
9: **end for**
10: Evaluate each candidate $\boldsymbol{\theta}_n^{(i)}$ on its assigned target task $\mathcal{T}_{\tau_n^{(i)}}$
11: Update the corresponding archives $\{\mathcal{A}_n\}_{n=1}^N$ with the evaluated candidates
12: Evolve the Gaussian distributions maintained in $\{e_n\}_{n=1}^N$ using the combined feedback from QD performance and subspace projected repulsion as introduced in **Section 4.2**
13: Update the mean policies through the PPO updates induced by the evolved PPGA emitters
14: Update $\mathcal{P}$ according to the adaptation method introduced in **Section 4.3**
15: **end for**
16: **return** Archives $\{\mathcal{A}_n\}_{n=1}^N$

update directions estimated through multiple PPO updates. Candidate policies are then generated by combining these directions with coefficients sampled from the xNES distribution, resulting in an arborescent search around the current policy. After updating the coefficient distribution according to QD performance and the proposed subspace-projected repulsion, its mean coefficients are used to determine a weighted combination of the objective and behaviors, and the search point of the PPGA emitter is subsequently to optimize this weighted combination through multiple PPO updates, following the mean-update procedure of PPGA. The cross-task transfer and its adaptive probability mechanism are otherwise applied in the same manner as in the standard MTQD-CoD framework.

D.2 Parameter Settings

We configure MTQD-CoD separately for the three problem domains according to their underlying differentiable QD backbones. For a controlled comparison, MTQD-CoD retains the main emitter, archive, and task-optimization settings of the corresponding baselines, i.e., CMA-MEGA, CMA-MAEGA, or PPGA, and introduces cross-task collaboration through subspace projected repulsion and inter-task transfer probabilities adaptation. For the arm repertoire problems and the latent space illumination problems, both the baseline algorithms and the proposed MTQD-CoD are implemented via the `Pyribs` (Tjanaka et al., 2023) package. For the policy optimization problems, we implement the algorithms based on the source code provided in `https://github.com/adaptive-intelligent-robotics/QDAC`. The detailed configurations are described below.

Arm Repertoire Problems. For the arm repertoire problems, we compare MTQD-CoD with OG-MAP-Elites, OMG-MEGA, CMA-MEGA, and CMA-MAEGA. All methods use the same $D = 1000$-dimensional decision space, a two-dimensional behavior descriptor over $[-1, 1]^2$, and a 100×100 grid archive. Each method is run for $T = 2000$ iterations over 20 independent seeds. To ensure a matched evaluation budget, each task performs one differentiable QD evaluation together with a batch of $B = 36$ candidate evaluations per iteration, resulting in 37 evaluations per task per iteration.

OG-MAP-Elites and OMG-MEGA use the `GradientOperatorEmitter` in `Pyribs` with perturbation scale $\sigma = 0.05$, gradient step size $\sigma_g = 0.1$, and batch size $B = 36$. Both use isotropic variation without gradient normalization. The main difference is that OMG-MEGA additionally uses behavior-measure gradients, whereas OG-MAP-Elites does not. Both methods use the standard archive configuration without annealing.

Table 7: Environment-specific hyperparameters for the policy optimization problems.

Parameter	Walker2d	Ant	Humanoid
Behavior Dim.	2	4	2
Archive Bins Per Dim.	50	10	50
Archive Learning Rate	0.5	0.1	0.1
`threshold_min`	-10	-500	200
σ_0	1.0	3.0	0.5
PPO Learning Rate	1×10^{-3}	1×10^{-3}	3×10^{-4}
Observation/Reward Clipping	No	No	Yes

CMA-MEGA and CMA-MAEGA use the `GradientArborescenceEmitter` with initial coefficient-distribution scale $\sigma_0 = 0.05$, solution-point learning rate 1.0, and batch size $B = 36$. The two methods use the same emitter configuration and differ only in the archive update. CMA-MEGA uses a standard archive, whereas CMA-MAEGA uses an annealing archive with learning rate 0.01.

For MTQD-CoD, we construct variants on top of both CMA-MEGA and CMA-MAEGA. In each case, the corresponding emitter, batch size, and archive configuration are kept identical to the backbone, while adaptive inter-task transfer and subspace projected repulsion are enabled. Specifically, the initial cross-task transfer probability is set to 0.3. We set the softmax temperature $\tau = 0.05$, transfer adaptation learning rate to $\eta = 0.1$, the exponential reward decay factor to $\rho = 0.9$, and the independence-bias coefficient to $\beta = 0.1$. For subspace projected repulsion, the objective gradient is included in the local differentiable search subspace, and the repulsion weight α is scheduled from 0 to 0.3 using a cubic schedule.

For archive admission, all archives use the admission rule of pyribs' `GridArchive`. We consider two settings: a greedy archive, which retains the best-so-far solution in each cell, and a annealing MAE-style archive, which allows slightly worse candidates to be admitted through a lagging threshold and thus provides smoother archive updates. OG-MAP-Elites, OMG-MEGA, and *MTQD-CoD (ME)* use the greedy setting, whereas CMA-MAEGA and *MTQD-CoD (MAE)* use the soft-threshold setting. The final reported archives are evaluated using the greedy archive.

Latent Space Illumination Problems. For the latent space illumination problems, we compare MTQD-CoD with OG-MAP-Elites, OMG-MEGA, CMA-MEGA, and CMA-MAEGA. All methods optimize the same 7168-dimensional normalized StyleGAN2 latent space and use a batch size of $B = 36$. No explicit clipping is applied to the latent decision variables.

OG-MAP-Elites and OMG-MEGA use a perturbation scale of $\sigma = 0.1$, a gradient step size of $\sigma_g = 0.002$, and a batch size of $B = 36$. CMA-MEGA and CMA-MAEGA use $\sigma_0 = 0.002$, a solution-point learning rate of 1.0, and a batch size of $B = 36$. CMA-MEGA uses the standard archive configuration, while CMA-MAEGA uses the annealing archive configuration. MTQD-CoD follows the corresponding CMA-MEGA or CMA-MAEGA backbone configuration and additionally enables adaptive inter-task transfer and subspace projected repulsion. The initial cross-task transfer probability is set to 0.3. We set the softmax temperature $\tau = 0.05$, the transfer adaptation learning rate to $\eta = 0.1$, the exponential reward decay factor to $\rho = 0.9$, and the independence-bias coefficient to $\beta = 0.1$. For subspace projected repulsion, the objective gradient is included in the local differentiable search subspace, and the repulsion weight α is scheduled from 0 to 0.35 using a cubic schedule. All remaining settings are kept identical to the corresponding backbone. For archive admission, we following the same setting with the arm repertoire problems.

Policy Optimization Problems. For policy optimization, we compare MTQD-CoD with PPGA, QD-PG, and DCG-ME. The hyperparameter settings of QD-PG and DCG-ME follow those reported in QDAC (Grillotti et al., 2024). All methods are run for $T = 5000$ iterations. For PPGA and MTQD-CoD, we use an environment batch size of 3000, a candidate population size of $B = 300$, and a rollout length of 128. PPO uses a discount factor of $\gamma = 0.99$, a GAE parameter of $\lambda_{\text{GAE}} = 0.95$, 4 update epochs, 8 minibatches, a clipping coefficient of 0.2, an entropy coefficient of 0, a value-loss coefficient of 2.0, and a maximum gradient norm of 1.0. Both the gradient-estimation stage and the emitter-mean update stage perform 10 PPO iterations. Observations and returns are normalized throughout training.

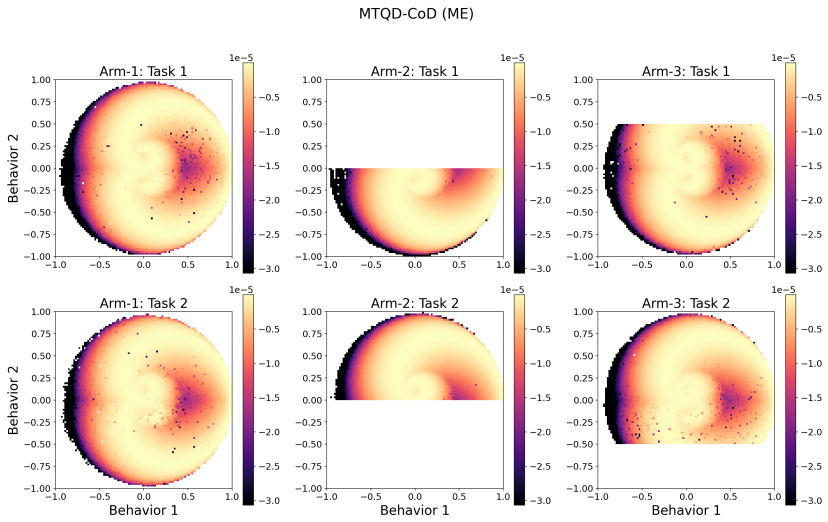

Figure 7: Heatmap visualizations of MTQD-CoD (ME) in the behavior spaces of the arm repertoire problems. We use 100×100 archives for clearer visualization of the discovered solutions.

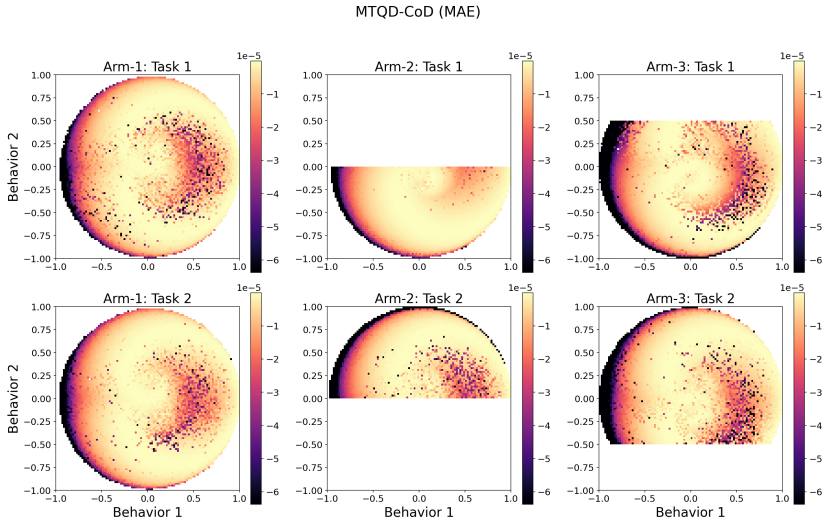

Figure 8: Heatmap visualizations of MTQD-CoD (MAE) in the behavior spaces of the arm repertoire problems. We use 100×100 archives for clearer visualization of the discovered solutions.

The environment-specific configurations are summarized in Table 7. For Walker2d, we use $\sigma_0 = 1.0$, a PPO learning rate of 1×10^{-3}, and a 50×50 archive with learning rate 0.5 and $\texttt{threshold_min} = -10$ of archive admission. For Ant, we use $\sigma_0 = 3.0$, a PPO learning rate of 1×10^{-3}, and a four-dimensional archive with 10 bins per behavior dimension, together with an archive learning rate of 0.1 and $\texttt{threshold_min} = -500$. For Humanoid, we use $\sigma_0 = 0.5$, a PPO learning rate of 3×10^{-4}, and a 50×50 archive with learning rate 0.1 and $\texttt{threshold_min} = 200$. Observation and reward clipping is enabled only for Humanoid.

MTQD-CoD follows the corresponding PPGA configuration and additionally enables adaptive inter-task transfer and subspace projected repulsion. The initial cross-task transfer probability is set to 0.3. We set the softmax temperature $\tau = 0.05$, the transfer adaptation learning rate to $\eta = 0.1$, the exponential reward decay factor to $\rho = 0.9$, and the independence-bias coefficient to $\beta = 0.1$. For subspace projected repulsion, the repulsion weight α is scheduled from 0 to 0.25, the ridge coefficient is set to $\lambda = 10^{-6}$, and the covariance stabilization term is set to $\varepsilon = 10^{-3}$. All remaining PPGA, PPO, and archive settings are kept unchanged.

E ADDITIONAL RESULTS ON ARM REPERTOIRE PROBLEMS.

In this section, we further visualize the behavior-space distributions obtained by MTQD-CoD (ME) and MTQD-CoD (MAE) using heatmaps in Fig. 7 and Fig. 8. On Arm-1, both variants achieve broad coverage of the reachable behavior space, with only a small number of cells remaining unoccupied. On Arm-2 and Arm-3, where task-specific obstacles restrict the accessible behaviors, both variants populate the remaining accessible regions while largely avoiding the blocked areas. These visualizations suggest that MTQD-CoD can adapt its search to different arm configurations and obstacle settings. A qualitative difference can also be observed between the two archive configurations. MTQD-CoD (ME) tends to exhibit higher-quality solutions within the behavior cells it discovers, whereas MTQD-CoD (MAE) tends to achieve broader behavior-space coverage. This distinction is not always visually pronounced in the heatmaps and should therefore be interpreted together with the quantitative results in Table 1, where the relative differences in coverage and normalized QD score are reported more explicitly.

F ADDITIONAL RESULTS ON LATENT SPACE ILLUMINATION PROBLEMS

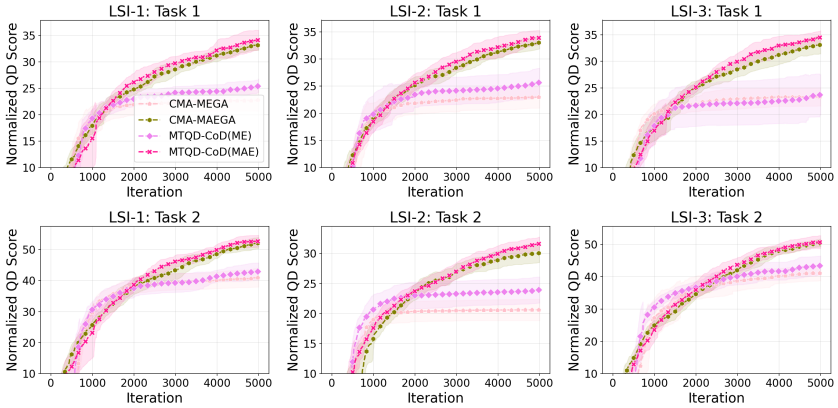

Figure 9: Comparison of normalized QD score convergence trends averaged over 10 independent runs of CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE) on latent-space illumination problems. Shaded areas represent one standard deviation on either side of the mean.

Here, we provide additional results on the convergence of MTQD-CoD on the latent space illumination problems. Fig. 9 reports the convergence trends of normalized QD score throughout optimization for CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE). Consistent with the coverage trends, MTQD-CoD (ME) converges faster than CMA-MEGA on both LSI-2 tasks and remains broadly comparable on the other tasks. MTQD-CoD (MAE) also shows modestly faster convergence than CMA-MAEGA on LSI-2 Task 2 and LSI-3 Task 1, while exhibiting similar behavior elsewhere. In Fig. 10, we additionally visualize representative image collages generated by MTQD-CoD in the LSI (StyleGAN) domain, showing diverse samples that vary along the considered behavioral features. Finally, Fig. 11, 12, and 13 present example behavior-space archives produced by the four methods. These visualizations are consistent with the quantitative results, with the MTQD-CoD variants tending to occupy broader portions of the behavior space than their corresponding backbones.

G ADDITIONAL RESULTS ON POLICY OPTIMIZATION PROBLEMS

We further show the additional convergence behavior of MTQD-CoD and its backbone PPGA on the policy optimization tasks. Specifically, Fig. 14 and Fig. 15 report the evolution of coverage and normalized QD score throughout optimization, respectively. Across the three environments, MTQD-CoD generally exhibits faster convergence than PPGA on both metrics, indicating that cross-task collaborative discovery can help the underlying policy-QD search reach broader and higher-quality



Figure 10: Latent space illumination collages generated by MTQD-CoD across three two-task problems.

archives with fewer optimization iterations. The improvement is particularly noticeable on Humanoid, where the gap between MTQD-CoD and PPGA becomes more evident during optimization. Since MTQD-CoD is built directly on PPGA and retains its basic policy-search framework, the observed gains are consistent with the additional cross-task coordination introduced by collaborative discovery rather than a change in the underlying policy optimizer itself. To further illustrate the resulting archive structures, Fig. 16 visualizes representative behavior-space heatmaps obtained by MTQD-CoD and PPGA. The heatmaps show that MTQD-CoD tends to occupy broader regions of the behavior space, which is consistent with the coverage convergence results and provides additional qualitative evidence that collaborative discovery can improve the exploration of behaviorally distinct policies.

H Ablation Studies

In MTQD-CoD, we introduce two mechanisms for cross-task collaborative discovery: subspace projected repulsion and inter-task transfer probability adaptation. To examine their individual contributions, we consider two ablated variants: MTQD-CoD-w/o-R, which removes the repulsion mechanism, and MTQD-CoD-w/o-T, which disables cross-task transfer by setting all cross-task transfer probabilities to zero. We evaluate these variants with both the standard ME archive and the MAE archive on the arm repertoire and latent space illumination problems. The corresponding coverage

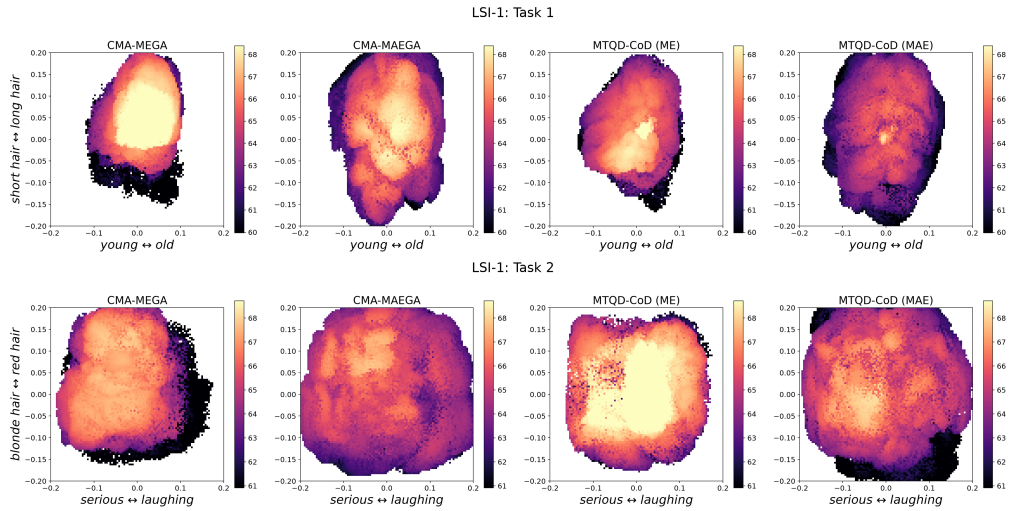

Figure 11: Heatmap visualizations of CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE) in the behavior spaces of the two tasks of LSI-1. We use 50×50 archives for clearer visualization of the discovered solutions.

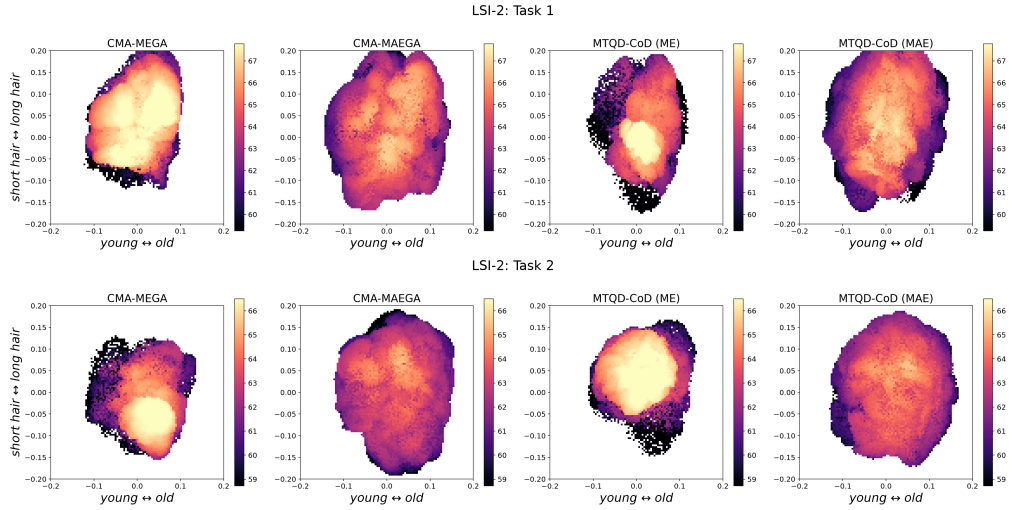

Figure 12: Heatmap visualizations of CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE) in the behavior spaces of the two tasks of LSI-2. We use 50×50 archives for clearer visualization of the discovered solutions.

and normalized QD score results are reported in Table 8. Overall, the full MTQD-CoD variants achieve favorable aggregate performance compared with their corresponding ablations, while the relative contribution of repulsion and transfer varies across different task relationships.

For the arm repertoire problems, this variation is reflected by the different levels of relatedness between the two tasks. In Arm-1, the two tasks differ only in link lengths and contain no task-specific obstacles, leading to highly similar behavior spaces and target solution regions between the tasks. Under this setting, both MTQD-CoD (ME) and MTQD-CoD (MAE) outperform their corresponding ablated variants, suggesting that repulsion and transfer can be jointly beneficial. In Arm-2, the two tasks use opposing obstacle configurations and therefore favor largely complementary regions of the behavior space. Here, the MTQD-CoD-w/o-T variants remain particularly competitive and can outperform the full variants, suggesting that subspace projected repulsion is more important in this setting while inter-task transfer provides a smaller benefit. Arm-3 presents an intermediate case, where the two tasks share a central accessible region while retaining task-specific differences, and the full MTQD-CoD variants achieve the strongest overall performance among the corresponding

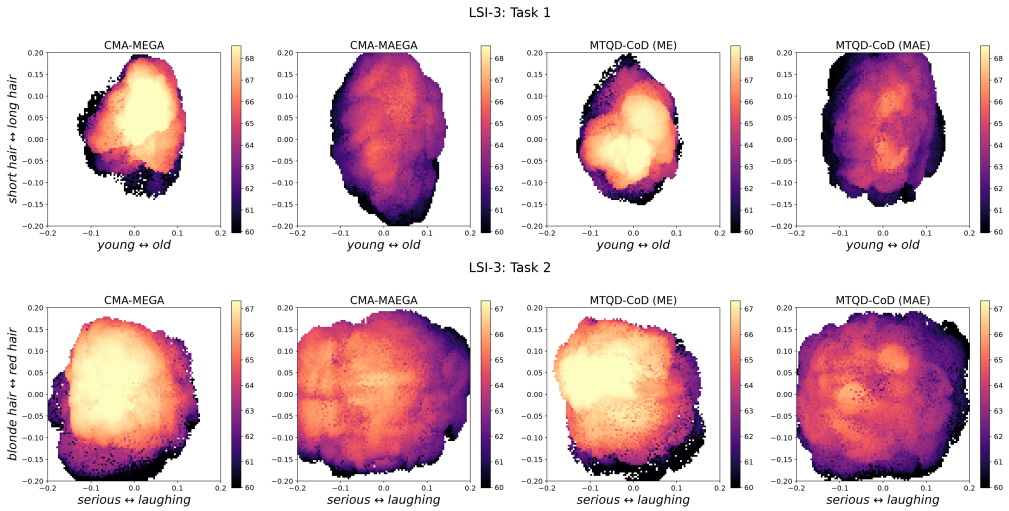

Figure 13: Heatmap visualizations of CMA-MEGA, CMA-MAEGA, MTQD-CoD (ME), and MTQD-CoD (MAE) in the behavior spaces of the two tasks of LSI-3. We use 50×50 archives for clearer visualization of the discovered solutions.

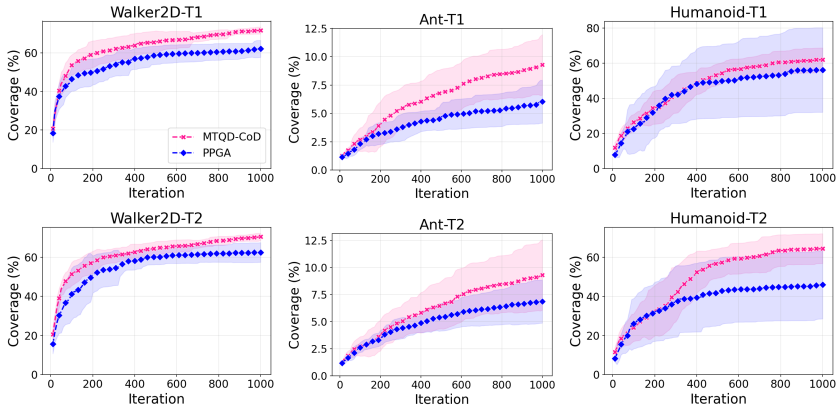

Figure 14: Comparison of coverage convergence trends averaged over 5 independent runs of PPGA and MTQD-CoD on policy optimization problems. Shaded areas represent one standard deviation on either side of the mean.

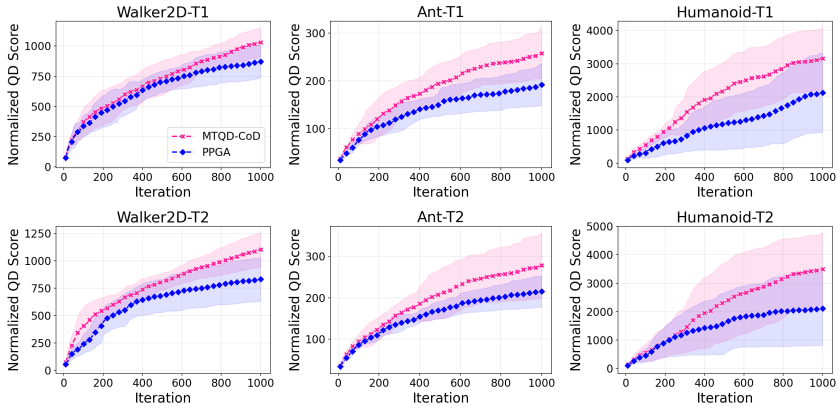

Figure 15: Comparison of normalized QD score convergence trends averaged over 5 independent runs of PPGA and MTQD-CoD on policy optimization problems. Shaded areas represent one standard deviation on either side of the mean.

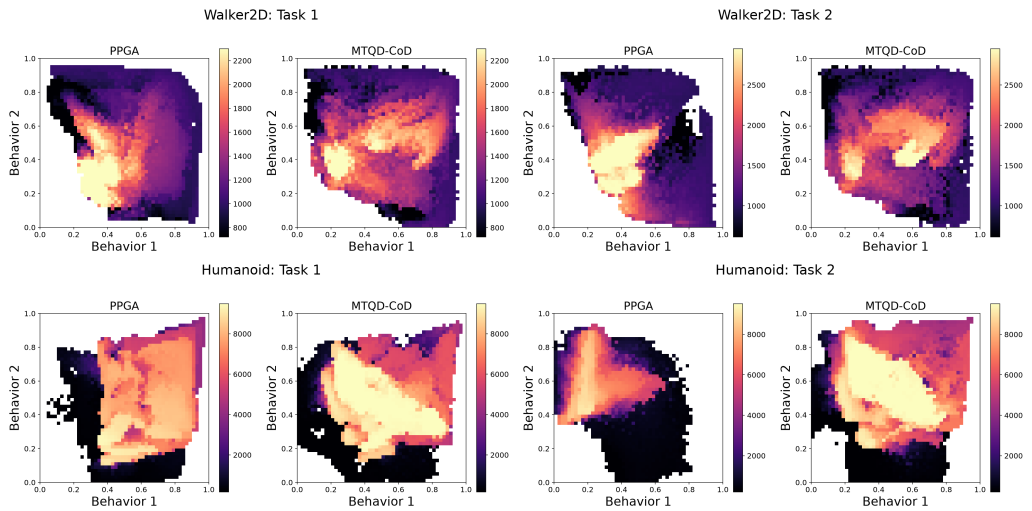

Figure 16: Heatmap visualizations of the solutions discovered by PPGA and MTQD-CoD in the behavior spaces of the Walker2D- and Humanoid-based two-task policy optimization problems. We use 50×50 archives for clearer visualization.

Table 8: Ablation Studies on the effectiveness of the subspace projected repulsion and adaptive inter-task transfer in MTQD-CoD.

Problem / Task	MTQD-CoD (MAE)	MTQD-CoD-w/o-R (MAE)	MTQD-CoD-w/o-T (MAE)	MTQD-CoD (ME)	MTQD-CoD-w/o-R (ME)	MTQD-CoD-w/o-T (ME)
Coverage						
Arm-1 / Task 0	$\mathbf{75.96\% \pm 0.49\%}$	$75.69\% \pm 0.39\%$	$75.73\% \pm 0.62\%$	$\mathbf{74.54\% \pm 0.72\%}$	$73.71\% \pm 0.79\%$	$73.89\% \pm 0.87\%$
Arm-1 / Task 1	$75.91\% \pm 0.60\%$	$\mathbf{75.93\% \pm 0.43\%}$	$75.53\% \pm 0.57\%$	$\mathbf{74.48\% \pm 0.77\%}$	$73.63\% \pm 0.88\%$	$73.50\% \pm 1.01\%$
Arm-2 / Task 0	$\mathbf{38.92\% \pm 0.67\%}$	$38.45\% \pm 0.42\%$	$38.73\% \pm 0.62\%$	$37.59\% \pm 0.91\%$	$37.15\% \pm 0.64\%$	$\mathbf{37.79\% \pm 0.94\%}$
Arm-2 / Task 1	$38.72\% \pm 0.58\%$	$38.71\% \pm 0.55\%$	$\mathbf{38.75\% \pm 0.62\%}$	$37.47\% \pm 1.10\%$	$37.05\% \pm 0.72\%$	$\mathbf{37.54\% \pm 1.02\%}$
Arm-3 / Task 0	$\mathbf{61.79\% \pm 0.91\%}$	$61.65\% \pm 0.35\%$	$61.51\% \pm 0.63\%$	$\mathbf{60.40\% \pm 0.90\%}$	$59.49\% \pm 0.84\%$	$59.74\% \pm 0.96\%$
Arm-3 / Task 1	$61.67\% \pm 0.71\%$	$\mathbf{61.68\% \pm 0.64\%}$	$61.62\% \pm 0.61\%$	$\mathbf{60.21\% \pm 0.54\%}$	$59.46\% \pm 0.80\%$	$60.34\% \pm 0.84\%$
LSI-1 / Task 1	$53.67\% \pm 3.29\%$	$50.55\% \pm 2.38\%$	$\mathbf{53.83\% \pm 2.25\%}$	$\mathbf{39.54\% \pm 1.99\%}$	$37.88\% \pm 2.94\%$	$38.99\% \pm 3.98\%$
LSI-1 / Task 2	$\mathbf{81.70\% \pm 2.86\%}$	$77.25\% \pm 3.47\%$	$81.53\% \pm 3.37\%$	$65.52\% \pm 4.36\%$	$62.12\% \pm 5.38\%$	$\mathbf{68.16\% \pm 4.07\%}$
LSI-2 / Task 1	$53.33\% \pm 2.27\%$	$51.02\% \pm 1.59\%$	$\mathbf{54.06\% \pm 2.77\%}$	$\mathbf{41.61\% \pm 3.31\%}$	$39.34\% \pm 2.65\%$	$37.90\% \pm 3.74\%$
LSI-2 / Task 2	$50.50\% \pm 1.73\%$	$48.49\% \pm 1.91\%$	$\mathbf{50.57\% \pm 2.37\%}$	$\mathbf{38.49\% \pm 3.00\%}$	$37.58\% \pm 1.58\%$	$36.83\% \pm 2.03\%$
LSI-3 / Task 1	$\mathbf{54.42\% \pm 2.01\%}$	$50.15\% \pm 1.48\%$	$53.36\% \pm 0.78\%$	$\mathbf{38.28\% \pm 4.28\%}$	$36.99\% \pm 3.24\%$	$38.14\% \pm 2.13\%$
LSI-3 / Task 2	$\mathbf{80.06\% \pm 3.28\%}$	$77.25\% \pm 2.42\%$	$79.13\% \pm 2.25\%$	$67.84\% \pm 5.06\%$	$63.45\% \pm 3.65\%$	$\mathbf{69.93\% \pm 4.99\%}$
Normalized QD Score						
Arm-1 / Task 0	$\mathbf{3.80 \pm 0.02}$	3.78 ± 0.02	3.79 ± 0.03	$\mathbf{3.73 \pm 0.04}$	3.69 ± 0.04	3.69 ± 0.04
Arm-1 / Task 1	3.80 ± 0.03	$\mathbf{3.80 \pm 0.02}$	3.78 ± 0.03	$\mathbf{3.72 \pm 0.04}$	3.68 ± 0.04	3.68 ± 0.05
Arm-2 / Task 0	$\mathbf{1.95 \pm 0.03}$	1.92 ± 0.02	1.94 ± 0.03	1.88 ± 0.05	1.86 ± 0.03	$\mathbf{1.89 \pm 0.05}$
Arm-2 / Task 1	1.94 ± 0.03	1.94 ± 0.03	$\mathbf{1.94 \pm 0.03}$	1.87 ± 0.05	1.85 ± 0.04	$\mathbf{1.88 \pm 0.05}$
Arm-3 / Task 0	$\mathbf{3.09 \pm 0.05}$	3.08 ± 0.02	3.08 ± 0.03	$\mathbf{3.02 \pm 0.04}$	2.97 ± 0.04	2.99 ± 0.05
Arm-3 / Task 1	3.08 ± 0.04	$\mathbf{3.08 \pm 0.03}$	3.08 ± 0.03	$\mathbf{3.01 \pm 0.03}$	2.97 ± 0.04	3.02 ± 0.04
LSI 1 / Task 1	34.12 ± 2.02	32.44 ± 1.38	$\mathbf{34.26 \pm 1.39}$	$\mathbf{25.33 \pm 1.36}$	23.14 ± 2.92	24.75 ± 2.86
LSI 1 / Task 2	$\mathbf{52.66 \pm 2.02}$	50.47 ± 2.04	52.11 ± 2.16	42.94 ± 2.91	40.93 ± 3.54	$\mathbf{44.58 \pm 2.19}$
LSI 2 / Task 1	33.93 ± 1.39	32.87 ± 1.05	$\mathbf{34.32 \pm 1.79}$	$\mathbf{25.64 \pm 2.77}$	24.53 ± 2.24	24.18 ± 2.27
LSI 2 / Task 2	31.57 ± 1.09	30.55 ± 1.17	$\mathbf{31.64 \pm 1.50}$	$\mathbf{23.92 \pm 2.30}$	22.83 ± 1.58	23.26 ± 1.42
LSI 3 / Task 1	$\mathbf{34.46 \pm 1.24}$	32.17 ± 1.09	33.99 ± 0.66	23.60 ± 4.27	23.11 ± 3.02	$\mathbf{24.26 \pm 2.48}$
LSI 3 / Task 2	$\mathbf{50.74 \pm 2.06}$	49.45 ± 1.48	50.13 ± 1.37	$\mathbf{43.42 \pm 3.13}$	40.78 ± 2.52	41.22 ± 9.64

ablations. This interpretation is further supported by the transfer dynamics in Fig. 17. As shown in the figure, the transfer probabilities remain relatively high on Arm-1, decrease rapidly on Arm-2, and exhibit an intermediate pattern on Arm-3. These results indicate that the adaptive transfer mechanism responds to the degree of shared structure between tasks rather than enforcing a fixed level of transfer.

For the latent space illumination problems, both the ME and MAE versions of MTQD-CoD-w/o-T remain competitive, and in five cases achieve better coverage and normalized QD scores than their corresponding full MTQD-CoD variants. One possible explanation is that the paired latent space illumination tasks often share only limited useful regions in the latent space because their target identities, semantic behavior axes, or both can differ. In such cases, subspace projected repulsion can

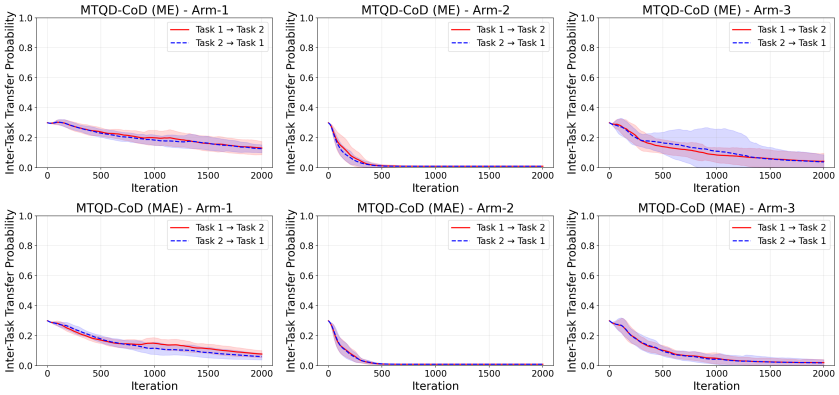

Figure 17: Dynamics of the cross-task transfer probabilities in MTQD-CoD (MAE) and MTQD-CoD (ME) across the three arm repertoire problems.

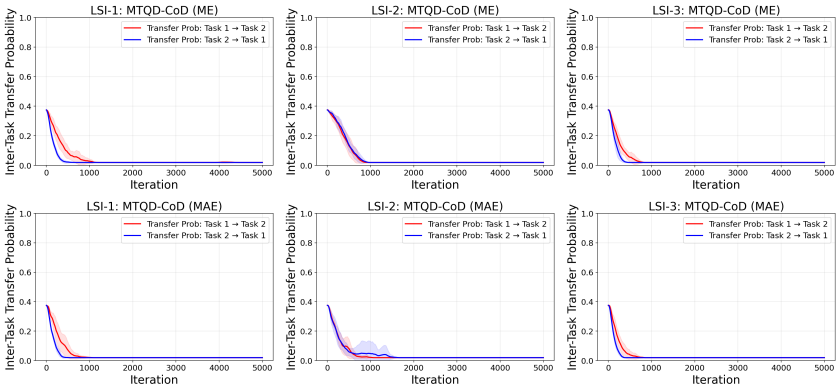

Figure 18: Dynamics of the cross-task transfer probabilities in MTQD-CoD (MAE) and MTQD-CoD (ME) across the three latent space illumination problems.

still encourage complementary exploration, whereas direct cross-task transfer may provide limited additional benefit. Consistent with this observation, the transfer probabilities in both directions decrease rapidly across the latent space illumination problems, as shown in Fig. 18, indicating that the adaptive mechanism progressively reduces transfer when transferred candidates contribute less to the target archives.