GIFT-EVO: FEEDBACK ADAPTATION THROUGH GEOMETRIC-INDUCTIVE OPERATORS FOR LOW-BUDGET BLACK-BOX OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Learned evolutionary optimizers encode cross-task strategies into fixed parameters via offline training, allowing direct search on new problems without per-task tuning for low-budget black-box optimization. However, existing methods do not retain real-time search feedback, preventing the policy from exploiting past outcomes to guide later decisions and accelerate convergence. Online parameter updates adapt to the target but require extensive samples, conflicting with low budgets. To this end, we propose GIFT-Evo, which introduces a lightweight search knowledge channel that accumulates individual acceptance and improvement history. This state modulates offspring generation to better suit the current problem, thereby accelerating convergence, while keeping parameters fixed. Since search knowledge must be translated through the model, and absolute-coordinate-based models fail under scale changes, we further design a geometric-inductive operator framework based on relative population geometry to ensure consistent semantics across problem scales. Experiments on multiple baselines show that GIFT-Evo achieves significant advantages in both low-budget and cross-scale deployment scenarios.

1 INTRODUCTION

Black-box optimization is prevalent in scenarios such as hyperparameter tuning Hutter et al. (2019), neural architecture search Ye et al. (2020), neuroevolution Such et al. (2017); Ye et al. (2020), experimental design, and simulation-based engineering design Forrester & Keane (2009). Its core challenge lies in the absence of analytical expressions for the objective function, together with the often-limited evaluation budgets encountered in practical applications. Evolutionary algorithms have emerged as important tools for black-box optimization because of their population-based search capabilities and gradient-free nature Storn & Price (1997). However, their performance relies heavily on manually designed search strategies, making it difficult to maintain consistent effectiveness across diverse optimization problems. This limitation is especially serious under a low evaluation budget, as ineffective search decisions waste the few available evaluations.

To reduce the dependence on manual design, learning-based evolutionary optimizers learn cross-task search strategies through offline training Li et al. (2025; 2024); Lange et al. (2023a;b). The learned strategies are stored in model parameters, which we refer to as *model knowledge*. During deployment, this model knowledge directly guides population search on new target problems without per-task tuning, thus requiring only a small number of evaluations on the new problem. This is particularly appealing for low-budget black-box optimization. However, existing methods usually use parent and offspring outcomes only for selection; they do not retain acceptance, relative improvement, and rejection from previous generations as inputs to later operator decisions. We refer to this target-specific historical feedback as *search knowledge* Ma et al. (2025). Without search knowledge, earlier evaluations cannot inform future decisions, limiting the performance achievable within a given evaluation budget.

Another line of work updates optimizer parameters online using target-task feedback Wang et al. (2026b;a). However, reliable parameter updates require many evaluated samples, and they also increase deployment cost. Under a strict evaluation budget, only limited feedback is available,

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.

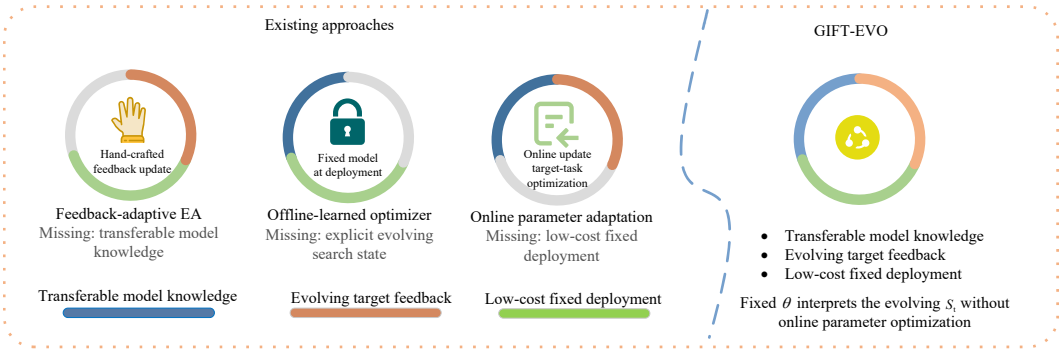


Figure 1: Motivation and high-level comparison of hand-designed evolutionary algorithms, offline meta-learned optimizers, and GIFT-Evo. GIFT-Evo combines transferable relative-structure model knowledge with evolving search-state feedback.

making online parameter optimization impractical. The central challenge is therefore to exploit limited evaluation feedback to improve later search decisions while keeping model knowledge fixed. Fig. 1 summarizes this challenge and shows how GIFT-Evo combines fixed model knowledge with evolving search knowledge.

GIFT-Evo addresses this challenge by keeping model knowledge fixed and introducing an independent search knowledge channel. Each individual maintains a state across generations. Its three channels record acceptance, relative improvement, and rejection. Parent and offspring outcomes update this state and guide later offspring generation. These outcomes are already available after selection, so no additional evaluations are required. To deploy the same model across dimensions, operator decisions should not depend on absolute coordinates. We therefore develop a geometric inductive operator framework using relative fitness rank, normalized fitness, and relative population geometry. These representations provide a consistent interface for the fixed model to translate evolving search knowledge into crossover, mutation, and fusion decisions. This design supports fixed parameter deployment from lower to higher dimensions. Fig. 2 and Section 3 detail the resulting optimization process.

The main contributions are summarized as follows:

- **Dual-knowledge evolutionary optimization paradigm.** We combine fixed, transferable model knowledge with an explicit and evolving search-knowledge channel, enabling feedback-based adaptation without extra evaluations or test-time parameter updates.
- **Geometric-inductive knowledge-alignment framework.** We use relative population geometry rather than absolute coordinates to map model knowledge and individual search states into scale-consistent crossover, mutation, and fusion behavior.
- **Controlled empirical evaluation.** We evaluate GIFT-Evo on 24 BBOB functions, planar-arm optimization, and Modified National Institute of Standards and Technology (MNIST) neuroevolution, including controlled low-budget comparisons, frozen cross-dimensional deployment, and task-specific training.

2 Related Work

2.1 Population-Based Algorithms

Population-based algorithms follow the same population search paradigm as our method. They maintain a population of candidate solutions and iteratively apply mutation, crossover, and selection to approach the optimum Holland (1962); Beyer & Schwefel (2002); Kennedy & Eberhart (1995); Storn & Price (1997). Their operator structures and adaptation rules are mainly designed by experts. Some methods use search outcomes to adjust control parameters Brest et al. (2006); Hansen & Ostermeier (1996; 2001); DaCosta et al. (2008). However, the rules that map these outcomes

to subsequent search decisions are still predefined. Therefore, their feedback adaptation remains dependent on manually designed mechanisms.

2.2 META-LEARNED ALGORITHMS

Meta-learned algorithms share with our method the core idea of learning search strategies from data. They can be further divided into the following two categories.

Offline learning methods follow the same train and deploy separation as our method. They encode search strategies into fixed model parameters that directly guide search on new problems Lange et al. (2023b;a); Li et al. (2024); Tian et al. (2025). During deployment, their decisions can respond to the current population and fitness. However, existing methods generally use parent and offspring outcomes for population selection without retaining individual acceptance, relative improvement, and rejection as cross-generation history. Consequently, these outcomes cannot directly guide later operator decisions.

Online learning methods use feedback from the target problem to adjust the search policy Wang et al. (2026b;a). Unlike offline methods, they update model parameters during deployment to adapt the policy to the current task. However, reliable parameter updates require sufficient evaluated samples and repeated optimization steps. Under a limited evaluation budget, the available feedback may be insufficient for the updated policy to take effect before the budget is exhausted. Therefore, these methods couple target-specific adaptation with online parameter optimization.

3 FRAMEWORK OF GIFT-EVO

3.1 PROBLEM SETUP AND NOTATION

We consider bounded black-box minimization:

$$\min_{x\in[\ell,u]^d} f(x), \tag{1}$$

where $x \in \mathbb{R}^d$ is the decision vector, $f(x)$ is accessible only through function evaluations, and B is the evaluation budget.

At generation t, the population and its fitness values are

$$X_t = [x_1^t, \dots, x_N^t]^\top \in \mathbb{R}^{N\times d}, \qquad F_t = [f(x_1^t), \dots, f(x_N^t)]^\top.$$

Each individual i also maintains a state

$$s_i^t = [s_{i,\mathrm{acc}}^t, s_{i,\mathrm{imp}}^t, s_{i,\mathrm{rej}}^t],$$

and $S_t = [s_1^t, \dots, s_N^t]^\top \in \mathbb{R}^{N\times 3}$ collects all individual states. We initialize $S_0 = \mathbf{0}$.

3.2 DUAL-KNOWLEDGE FRAMEWORK AND ONE-GENERATION ROLLOUT

GIFT-Evo separates two roles in candidate generation. **Search knowledge** is stored in S_t and summarizes the recent acceptance, improvement, and rejection of each individual. **Model knowledge** is encoded in the offline-trained parameters θ. It determines how the current population and search states are converted into crossover, mutation, and fusion decisions. During deployment, S_t is updated from target-task feedback, while θ remains fixed.

The model receives a relative context

$$R_t = \mathcal{R}(X_t, F_t),$$

which contains normalized fitness rank r_i^t, normalized fitness $\tilde{F}_i^t$, and relative population geometry such as the relative radius ρ_i^t. The learned modules predict population weights and scalar gates from these relative quantities. Their parameter shapes are independent of the decision dimension d. This design reduces dependence on objective magnitude, coordinate scale, and training dimension. Definitions of the relative features are given in Appendix A.1.

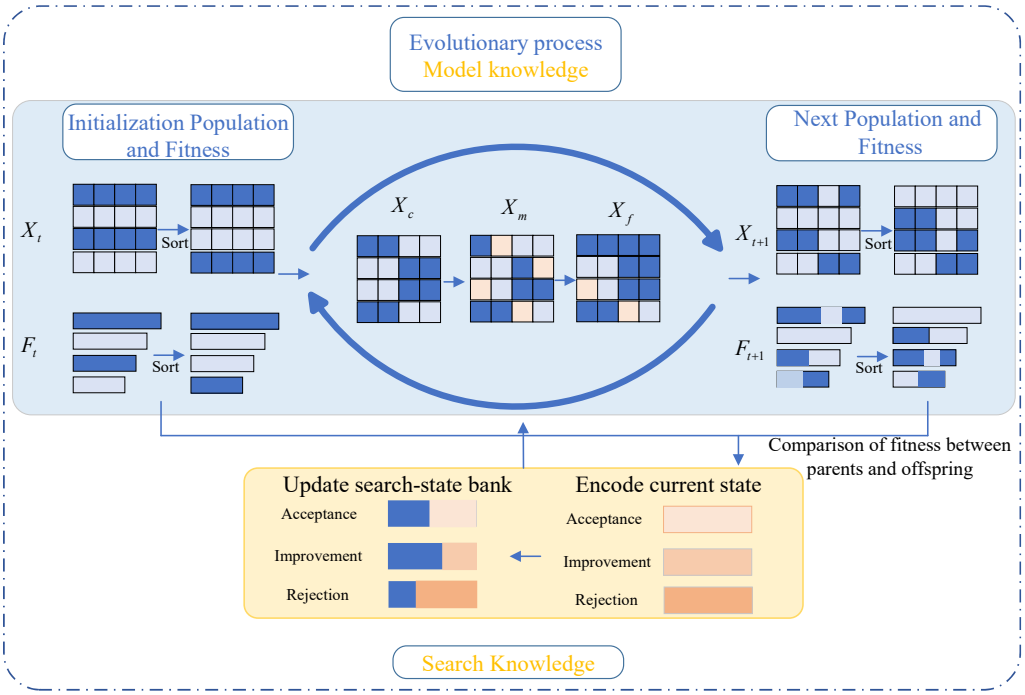


Figure 2: Framework of GIFT-Evo. Model knowledge governs the within-generation evolutionary process, including crossover, mutation, and fusion, whereas search knowledge is updated online from parent–offspring comparisons and retained across generations to guide subsequent evolution.

Fig. 2 shows one generation. GIFT-Evo jointly sorts (X_t, F_t, S_t), constructs R_t, and generates crossover and mutation candidates. Fusion produces N children, which are evaluated against their parents. Greedy selection yields (X_{t+1}, F_{t+1}), while the comparison outcomes update the search-state bank.

For individual i, the selection result produces an acceptance indicator a_i^t and a relative improvement δ_i^t:

$$a_i^t = \mathbb{I}\left[F_{\text{child},i}^t \le F_i^t\right], \qquad \delta_i^t = \frac{\max(0, F_i^t - F_{\text{child},i}^t)}{|F_i^t| + \epsilon}. \tag{2}$$

The current outcome is encoded and added to the state history through an exponential moving average:

$$\hat{s}_i^t = [a_i^t, \tanh(\delta_i^t), 1 - a_i^t], \qquad s_i^{t+1} = \beta s_i^t + (1-\beta)\hat{s}_i^t. \tag{3}$$

The first and third channels distinguish acceptance from rejection. The middle channel records the improvement magnitude. The updated state remains attached to the individual retained by selection and is used in the next generation. Appendix A.5 provides the complete update procedure.

3.3 State-Guided Operators with Geometric Inductive Bias

The three operators use R_t and S_t for different decisions. Crossover determines source weights, mutation controls geometric displacement, and fusion balances the parent, crossover, and mutation branches.

3.3.1 State-Guided Crossover

Crossover generates one population mixture for each target individual:

$$X_{\text{cross}}^t = A_t X_t, \qquad A_{ij}^t \ge 0, \qquad \sum_{j=1}^{N} A_{ij}^t = 1. \tag{4}$$

Row i of A_t contains the source weights used to construct the crossover candidate for individual i. The corresponding pairwise score is

$$e_{ij}^t = e_{ij}^{\text{rank}}(r_i^t, r_j^t) + \tanh(\alpha)\, e_{ij}^{\text{state}}(\tilde{F}_i^t, \tilde{F}_j^t, s_i^t, s_j^t) + b_{\text{self}}\mathbb{I}[i=j], \qquad A_{ij}^t = \frac{\exp(e_{ij}^t)}{\sum_{k=1}^N \exp(e_{ik}^t)}. \tag{5}$$

The rank term provides the base source preference. The state term allows the model to adjust this preference according to recent individual outcomes. The self-bias controls how much information the target individual retains. All three components are learned jointly during offline training.

3.3.2 State-Guided Mutation

Crossover remains within population mixtures. Mutation adds an exploratory displacement:

$$X_{\text{mut}}^t = X_{\text{cross}}^t + D_t \odot \Delta_t + U_t, \qquad D_i^t = \eta_i^t \lambda_i^t. \tag{6}$$

Here, Δ_t is an exploration direction synthesized from centered population differences. The predicted factor η_i^t and local-scale gate λ_i^t control the displacement magnitude. The bounded correction U_t moves the candidate toward a learned population reference when refinement is needed.

The state s_i^t enters the prediction of the direction weights, step scale, and correction gate. The learned model therefore decides how recent outcomes should change the movement of each individual. The displacement itself is constructed from population differences, so the learned parameters do not produce coordinate-specific mutation vectors.

3.3.3 State-Guided Fusion

Fusion combines three candidate branches:

$$x_{\text{child},i}^t = c_{p,i}^t x_i^t + c_{c,i}^t x_{\text{cross},i}^t + c_{m,i}^t x_{\text{mut},i}^t. \tag{7}$$

The three coefficients control parent preservation, population mixing, and exploratory movement. They are generated for each individual by

$$c_i^t = c_{\text{base},i} + \gamma \tanh\big(\Phi_\theta(r_i^t, s_i^t, \rho_i^t)\big), \qquad c_i^t = [c_{p,i}^t, c_{c,i}^t, c_{m,i}^t]. \tag{8}$$

The base vector describes the default branch balance learned offline. The state-conditioned term adjusts this balance during the target-task search. The coefficients are applied across all coordinates of the corresponding individual. They remain unconstrained, allowing the fusion operator to interpolate or extrapolate among the three branches.

Appendices A.2–A.4 give the complete network structures and operator parameterizations.

3.4 Offline Training and Fixed-Parameter Deployment

We train θ by unrolling GIFT-Evo on a set of proxy functions $\mathcal{F}_{\text{tr}}$. The objective measures the change in mean population fitness over the complete rollout:

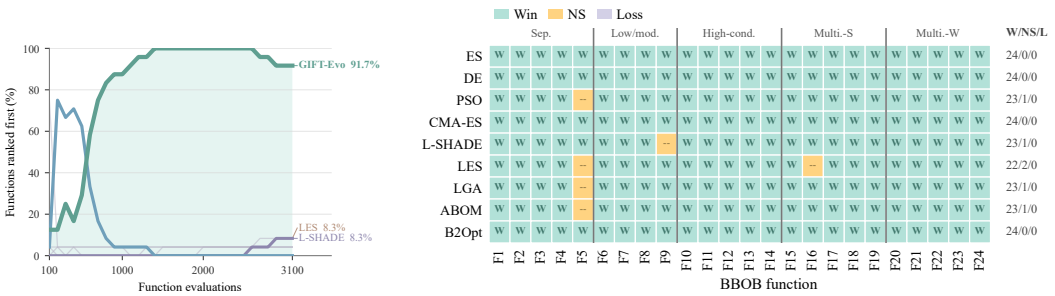
$$\mathcal{L}_{\text{off}} = \frac{1}{|\mathcal{F}_{\text{tr}}|} \sum_{f \in \mathcal{F}_{\text{tr}}} \left[\frac{\bar{f}(X_T) - \bar{f}(X_0)}{|\bar{f}(X_0)| + \epsilon} + \lambda_{\text{reg}} \mathcal{R}_{\text{score}} \right], \tag{9}$$

where

$$\bar{f}(X_t) = \frac{1}{N} \sum_{i=1}^N f(x_i^t).$$

The score regularizer $\mathcal{R}_{\text{score}}$ limits the state-conditioned crossover residual, and $\lambda_{\text{reg}} = 10^{-4}$ in all experiments. This objective trains the operators from population improvement and requires no expert trajectories.

At deployment, optimization begins with $S_0 = \mathbf{0}$. The trained θ is reused without parameter updates. Target-task evaluations influence later generations through the evolving state S_t. The training procedure, deployment algorithm, and computational complexity are reported in Appendices C, B, and D.



(a) Rank-1 coverage over the evaluation budget

(b) Holm-corrected function-level comparisons

Figure 3: Controlled comparison at $d = 10$ under the common 3,100-FE budget. (a) Percentage of BBOB functions on which each method attains the lowest median best-so-far error. Exact ties share Rank 1. (b) Holm-corrected win, non-significant-difference, and loss outcomes for GIFT-Evo against each baseline. Lower terminal error is better.

4 Experiments

We study four questions: RQ1 (**Controlled Low-Budget Performance**) compares GIFT-Evo with classical and learned optimizers under an identical budget; RQ2 (**Fixed-Parameter Cross-Dimensional Deployment**) tests whether a model trained at low dimension can be directly applied at higher dimensions without test-time updates; RQ3 (**Task-Specific Adaptability**) evaluates training the framework from scratch for robotic optimization and neuroevolution; and RQ4 (**Component Analysis**) examines the empirical contributions of search-state conditioning and the learned evolutionary operators.

4.1 Experimental Setup

Tasks GIFT-Evo is trained on CECF1–CECF3 at $d = 10$. RQ1 evaluates this model on the 24-function BBOB implementation bundled with B2Opt at $d = 10$; RQ2 directly deploys the same frozen checkpoint at $d = 30$ and $d = 100$. This is a BBOB implementation, rather than an official COCO benchmark evaluation. For RQ3, GIFT-Evo is trained separately on MNIST-based neuroevolution LeCun et al. (1998) and the planar mechanical-arm benchmark Cully et al. (2015). For RQ4, we evaluate the complete Full checkpoint and four deployment-time interventions on the held-out CECF4–CECF9 functions. All variants start from the same trained Full checkpoint, and no ablated variant is retrained. Task configurations and objectives are detailed in Appendix E.

Baselines We compare with (μ, λ)-ES Beyer & Schwefel (2002), DE/rand/1/bin Das & Suganthan (2010), PSO Kennedy & Eberhart (1995), CMA-ES Hansen (2016), L-SHADE Tanabe & Fukunaga (2014), LES Lange et al. (2023a), LGA Lange et al. (2023b), ABOM Wang et al. (2026a), and B2Opt Li et al. (2025). All BBOB methods receive the same total budget. GIFT-Evo uses one $d = 10$ checkpoint at every tested dimension, whereas B2Opt uses dimension-specific checkpoints; B2Opt is therefore a performance baseline, not a cross-dimensional-transfer comparator. MNIST and planar-arm B2Opt values are descriptive reference results.

Evaluation Protocol RQ1 and RQ2 use 30 runs (seeds $1, \ldots, 30$) and 3,100 function evaluations per method–function–dimension setting. We report terminal error as mean $\pm$ standard deviation. For each baseline and dimension, two-sided Mann–Whitney U tests on terminal errors are corrected across 24 functions using Holm correction at $\alpha = 0.05$; win/no-significant-difference/loss counts summarize the corrected tests. RQ3 uses five independent runs, whereas RQ4 uses 30. Both are interpreted descriptively without significance testing. Further implementation details are provided in Appendix E.

Table 1: Terminal BBOB error at $d = 10$ and the 3,100-FE reporting horizon. Entries are mean (standard deviation) over 30 independent runs; lower is better. Bold and underlined entries denote the best and second-best mean values. The final row gives Holm-corrected $\mathrm{W/NS/L}$ counts for GIFT-Evo against each baseline across 24 functions.

F	GIFT-Evo	ES	DE	PSO	CMA-ES	L-SHADE	LES	LGA	ABOM	B2Opt
F01	**2.406E-23 (4.068E-23)**	9.420E-01 (2.849E-01)	1.107E+00 (3.117E-01)	3.523E-02 (2.072E-02)	7.785E-03 (3.410E-03)	2.032E-05 (3.313E-05)	3.823E-03 (1.612E-03)	5.869E+00 (1.259E+01)	2.686E-01 (1.276E-01)	2.482E-04 (1.264E-04)
F02	**9.209E-24 (1.556E-23)**	2.458E-03 (9.117E-04)	1.261E-03 (4.336E-04)	1.869E-04 (1.029E-04)	4.020E-05 (2.913E-05)	9.383E-08 (1.440E-07)	1.501E-05 (1.071E-05)	3.387E-02 (3.239E-02)	1.335E-03 (6.942E-04)	1.042E-05 (5.731E-06)
F03	**2.206E-18 (4.747E-18)**	4.964E+01 (5.589E+00)	5.620E+01 (7.902E+00)	3.155E+01 (8.042E+00)	3.565E+01 (5.801E+00)	1.614E+01 (3.727E+00)	3.652E+01 (5.944E+00)	8.678E+01 (5.631E+01)	1.979E+01 (4.014E+00)	2.743E-01 (1.783E-01)
F04	**1.280E-11 (2.564E-11)**	6.353E+01 (6.909E+00)	7.503E+01 (7.587E+00)	4.195E+01 (9.030E+00)	4.830E+01 (6.669E+00)	2.406E+01 (5.174E+00)	4.123E+01 (6.793E+00)	2.195E+02 (9.491E+01)	3.508E+01 (8.830E+00)	5.969E+00 (3.706E+00)
F05	**0.000E+00 (0.000E+00)**	5.165E+00 (2.039E+00)	5.154E-01 (4.102E-01)	**0.000E+00 (0.000E+00)**	1.190E+01 (1.736E+00)	2.710E-06 (1.094E-05)	**0.000E+00 (0.000E+00)**	**0.000E+00 (0.000E+00)**	1.008E-02 (5.520E-02)	5.979E+00 (6.440E+00)
F06	**4.054E-19 (9.937E-19)**	2.648E+00 (6.246E-01)	6.994E+00 (2.572E+00)	1.757E-01 (8.596E-02)	3.335E-02 (1.958E-02)	3.688E-04 (2.962E-04)	3.861E-02 (1.249E-02)	3.362E+01 (2.644E+01)	2.073E+00 (8.748E-01)	1.993E-03 (1.045E-03)
F07	**2.467E-20 (2.288E-20)**	4.537E+00 (1.437E+00)	3.057E+01 (9.540E+00)	1.844E+00 (8.705E-01)	8.431E-02 (4.292E-02)	3.686E-02 (3.824E-02)	1.540E+00 (8.038E-01)	7.332E+01 (3.538E+01)	1.508E+01 (6.396E+00)	4.646E-09 (2.588E-09)
F08	**1.801E-16 (3.574E-16)**	9.580E+01 (2.485E+01)	2.292E+02 (7.438E+01)	1.675E+01 (1.540E+01)	1.027E+01 (1.178E+00)	7.455E+00 (1.131E+00)	1.137E+01 (1.282E+01)	4.088E+03 (9.641E+03)	1.530E+02 (5.536E+01)	9.684E-02 (6.236E-02)
F09	7.933E+00 (4.194E-01)	8.685E+01 (2.284E+01)	3.556E+02 (1.332E+02)	2.541E+01 (3.494E+01)	9.990E+00 (9.128E-01)	**7.702E+00 (1.019E+00)**	1.850E+01 (2.470E+01)	3.820E+03 (3.307E+03)	6.941E+01 (3.298E+01)	4.209E+01 (3.327E+00)
F10	**6.166E-15 (8.954E-15)**	8.387E+03 (4.443E+03)	5.179E+04 (2.150E+04)	3.237E+03 (2.011E+03)	1.016E+02 (6.689E+01)	1.156E+02 (3.122E+02)	6.474E+03 (5.743E+03)	6.748E+04 (5.250E+04)	2.920E+04 (1.341E+04)	1.466E+01 (1.166E+01)
F11	**8.784E-14 (1.378E-13)**	2.138E+01 (9.614E+00)	9.423E+01 (2.665E+01)	4.301E+01 (1.779E+01)	4.100E+00 (1.712E+00)	1.749E+00 (1.622E+00)	7.669E+01 (2.646E+01)	1.958E+02 (3.985E+01)	9.276E+01 (3.381E+01)	7.939E-01 (1.153E+00)
F12	**2.287E-13 (4.779E-13)**	6.478E+06 (1.937E+06)	1.768E+07 (4.967E+06)	2.326E+05 (1.163E+05)	7.824E+04 (5.848E+04)	2.809E+02 (3.595E+02)	6.219E+04 (3.920E+04)	1.499E+08 (1.663E+08)	3.582E+06 (1.555E+06)	2.207E+03 (1.103E+03)
F13	**2.778E-08 (2.722E-08)**	8.545E+01 (1.058E+01)	1.300E+02 (2.257E+01)	1.963E+01 (5.972E+00)	7.139E+00 (1.674E+00)	8.388E-01 (5.600E-01)	8.466E+00 (3.778E+00)	2.079E+03 (1.069E+02)	5.650E+01 (9.948E+00)	1.411E+00 (4.531E-01)
F14	**1.218E-11 (1.730E-11)**	5.045E-01 (1.023E-01)	1.940E+00 (5.176E-01)	3.735E-02 (1.380E-02)	1.608E-02 (5.111E-03)	8.765E-04 (6.603E-04)	1.542E-02 (4.360E-03)	1.055E+01 (4.568E+00)	4.528E-01 (3.116E-01)	2.620E-03 (1.394E-03)
F15	**2.185E-18 (8.792E-18)**	5.181E+01 (6.064E+00)	7.766E+01 (1.185E+01)	4.647E+01 (6.806E+00)	3.906E+01 (6.658E+00)	2.689E+01 (8.534E+00)	3.931E+01 (6.355E+00)	1.241E+02 (4.760E+01)	5.797E+01 (7.318E+00)	2.722E-01 (1.703E-01)
F16	4.257E+00 (1.592E+00)	1.346E+01 (3.075E+00)	1.333E+01 (2.590E+00)	7.346E+00 (3.326E+00)	1.265E+01 (3.010E+00)	8.070E+00 (3.418E+00)	**3.570E+00 (2.371E+00)**	9.793E+00 (4.339E+00)	1.292E+01 (3.205E+00)	1.542E+01 (3.613E+00)
F17	**3.060E-10 (3.907E-10)**	1.180E+00 (2.029E-01)	3.308E+00 (5.804E-01)	5.071E-01 (2.717E-01)	2.443E-01 (7.503E-02)	8.214E-02 (4.685E-02)	2.833E-01 (2.957E-01)	4.398E+00 (2.268E+00)	1.312E+00 (3.675E-01)	2.033E-02 (6.785E-03)
F18	**7.090E-10 (7.710E-10)**	4.101E+00 (6.057E-01)	1.156E+01 (2.332E+00)	2.105E+00 (6.458E-01)	8.446E-01 (1.692E-01)	2.540E-01 (1.225E-01)	1.365E+00 (1.125E+00)	1.802E+01 (1.097E+01)	4.744E+00 (1.129E+00)	7.385E-02 (2.569E-02)
F19	**5.957E-01 (2.100E-01)**	2.937E+00 (5.007E-01)	4.759E+00 (7.824E-01)	3.307E+00 (4.787E-01)	3.062E+00 (5.001E-01)	3.359E+00 (5.056E-01)	2.908E+00 (5.391E-01)	4.406E+00 (1.768E+00)	3.751E+00 (7.613E-01)	1.878E+00 (6.058E-01)
F20	**-2.253E-02 (1.028E+02)**	-4.668E+00 (5.461E+00)	-9.869E+00 (7.712E+00)	-7.002E+00 (6.781E+00)	-7.104E+00 (6.473E+00)	-7.920E+00 (6.389E+00)	6.172E+00 (4.845E+00)	2.471E+03 (2.995E+03)	-1.385E+01 (1.065E+01)	-1.163E+01 (9.013E+00)
F21	**2.227E-02 (3.814E-02)**	2.266E+00 (3.209E+00)	1.802E+01 (6.443E+00)	3.963E+00 (4.161E+00)	1.427E+00 (2.882E+00)	1.044E+00 (1.609E+00)	1.075E+01 (1.320E+01)	2.314E+01 (1.893E+01)	8.563E+00 (4.290E+00)	2.689E+01 (6.245E+00)
F22	**5.841E-01 (4.397E-01)**	1.199E+01 (4.735E+00)	2.000E+01 (5.695E+00)	7.251E+00 (7.682E+00)	8.489E+00 (2.580E+00)	3.706E+00 (3.767E+00)	1.316E+01 (1.135E+01)	4.251E+01 (1.916E+01)	1.077E+01 (3.236E+00)	4.840E+01 (5.746E+00)
F23	**1.066E+00 (1.836E-01)**	2.085E+00 (3.560E-01)	2.097E+00 (4.106E-01)	1.979E+00 (4.496E-01)	1.982E+00 (4.035E-01)	2.036E+00 (3.925E-01)	2.012E+00 (4.517E-01)	2.051E+00 (5.441E-01)	2.089E+00 (4.566E-01)	2.176E+00 (3.905E-01)
F24	**2.960E+01 (3.259E+00)**	5.613E+01 (5.931E+00)	8.121E+01 (9.391E+00)	5.422E+01 (8.803E+00)	4.651E+01 (6.764E+00)	4.870E+01 (8.236E+00)	4.563E+01 (7.076E+00)	8.738E+01 (3.018E+01)	6.346E+01 (7.332E+00)	8.145E+01 (8.851E+00)
GIFT-Evo W/NS/L		24/0/0	24/0/0	23/1/0	24/0/0	23/1/0	22/2/0	23/1/0	23/1/0	24/0/0

4.2 RQ1: Controlled Low-Budget Performance

RQ1 examines whether GIFT-Evo performs competitively under a common low evaluation budget. We evaluate the frozen checkpoint trained on CECF1–CECF3 at $d = 10$ on all 24 BBOB functions at the same dimension. Each method uses a population of 100, a maximum budget of 3,100 function evaluations (FEs), and 30 independent runs.

Table 1 reports the complete terminal results. The Holm-corrected $\mathrm{W/NS/L}$ outcomes follow the protocol in Appendix E.3. Terminal tables for $d = 30$ and $d = 100$ are omitted to avoid duplicating the common-horizon RQ2 evidence.

Fig. 3(a) shows the percentage of functions for which each method attains the lowest median best-so-far error as the budget increases. At 3,100 FEs, GIFT-Evo ranks first on 22 of the 24 functions, corresponding to a coverage of 91.7%. This advantage appears within the limited budget and remains through the later search stages.

We test terminal errors using two-sided Mann–Whitney U tests. For each baseline, Holm correction is applied across F1–F24. Fig. 3(b) reports the resulting function-level outcomes. Across the $9 \times 24 = 216$ comparisons, GIFT-Evo obtains 210 significant wins, 6 non-significant differences, and no significant losses. These counts summarize the pairwise tests and do not form an additional aggregate hypothesis test. Among the 210 wins, 209 have Holm-adjusted $p < 0.01$.

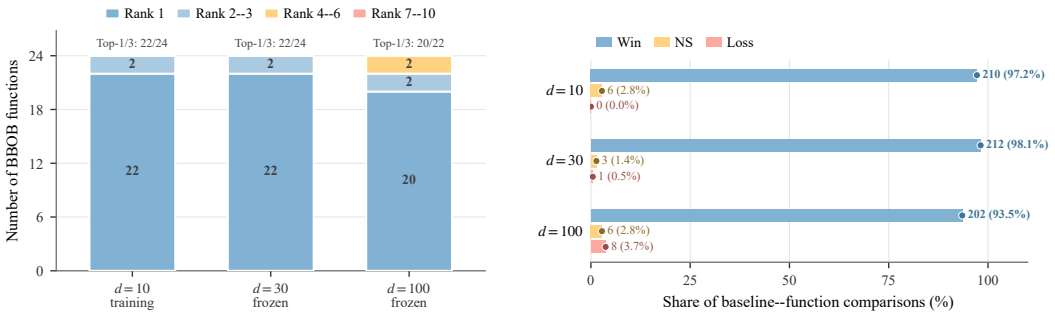
Together, these results show that GIFT-Evo performs strongly under the tested low-budget setting and has no significant losses against any of the nine baselines.

4.3 RQ2: Fixed-Parameter Cross-Dimensional Deployment

RQ2 evaluates whether the model trained at $d = 10$ can be deployed at higher dimensions without parameter updates. We apply the same frozen GIFT-Evo checkpoint at $d = 30$ and $d = 100$. High-dimensional BBOB data are excluded from training, hyperparameter selection, and checkpoint selection. The individual search states continue to evolve from online feedback, while the model parameters remain fixed. Every method receives 3,100 FEs in each setting.

Fig. 4(a) summarizes the terminal ranks of GIFT-Evo across the 24 functions. At $d = 10$, GIFT-Evo ranks first on 22 functions and is within the top three on all 24 functions. The corresponding counts are $22/24$ at $d = 30$ and $20/22$ at $d = 100$. Here, Top-1/3 denotes the number of functions ranked first and within the top three, respectively. The distribution therefore remains concentrated in the leading rank groups after dimensional transfer.

Fig. 4(b) reports the Holm-corrected comparisons against all nine baselines. GIFT-Evo obtains $210/6/0$ W/NS/L outcomes at $d = 10$, $212/3/1$ at $d = 30$, and $202/6/8$ at $d = 100$. Each dimension contains 216 baseline–function comparisons. Significant wins remain dominant after dimensional transfer, although the number of losses increases at $d = 100$.



(a) Terminal-rank distribution across dimensions

(b) Significance outcomes at each dimension

Figure 4: Cross-dimensional deployment of the GIFT-Evo checkpoint trained at $d = 10$. (a) Number of functions in each terminal-rank group. Top-1/3 gives the number of functions ranked first and within the top three, respectively. (b) Holm-corrected W/NS/L proportions over 216 baseline–function comparisons per dimension. All methods use the common 3,100-FE budget.

These results support direct deployment of the tested $d = 10$ checkpoint at $d = 30$ and $d = 100$. The conclusion is limited to the evaluated functions, dimensions, and budget. B2Opt uses dimension-specific checkpoints and serves as a performance reference in this experiment. Per-baseline outcomes are reported in Appendix F.

4.4 RQ3: Task-Specific Adaptability Beyond Synthetic Functions

RQ3 studies whether the GIFT-Evo framework can be trained for objectives outside the synthetic-function setting. We consider MNIST neuroevolution and planar mechanical-arm optimization. A task-specific GIFT-Evo model is trained for each experimental setting. These experiments evaluate adaptation through retraining. They do not measure zero-shot cross-task transfer. Each result uses five independent runs, so the comparisons are interpreted descriptively.

MNIST Neuroevolution. Fig. 5(a) reports test accuracy at four training-data fractions. GIFT-Evo achieves mean accuracies of 0.73, 0.79, 0.85, and 0.87. It has the highest reported mean at fractions 0.25 and 1.0, matches B2Opt at 0.75, and ranks second at 0.5. CMA-ES is the strongest remaining baseline across the four settings, with mean accuracies between 0.57 and 0.67. The separation from the conventional baselines is already visible when only one quarter of the training data is used.

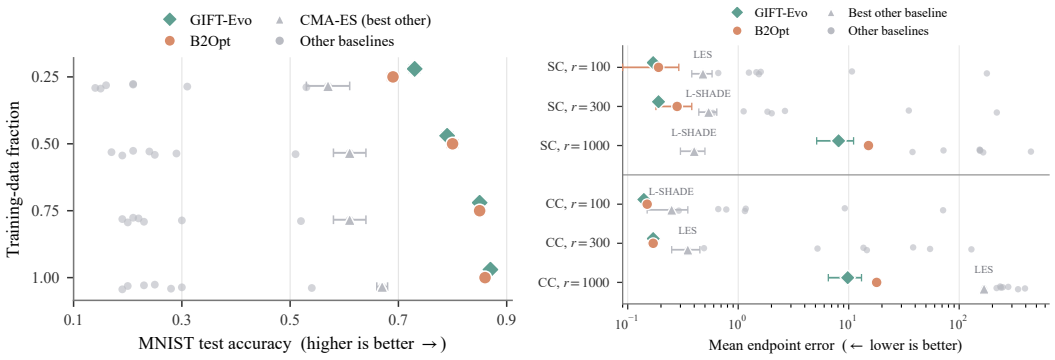
Planar Mechanical Arm. The planar-arm benchmark contains a simple case (SC) with fixed link lengths and a complex case (CC) that also optimizes link lengths. Figure 5(b) reports endpoint error for $r \in \{100, 300, 1000\}$. The logarithmic horizontal axis accommodates the large differences between methods.

In SC, GIFT-Evo obtains the lowest reported mean error at $r = 100$ and $r = 300$. At $r = 1000$, L-SHADE performs best and GIFT-Evo ranks second. In CC, GIFT-Evo has the lowest mean at $r = 100$ and $r = 1000$, and shares the lowest reported mean with B2Opt at $r = 300$. The figure highlights GIFT-Evo and B2Opt while retaining the remaining baselines as contextual references.

Taken together, the two tasks show that the framework can be trained for neural-network parameter optimization and robotic design. This evidence concerns task-specific training under the reported protocols. Statistical significance is not claimed from the five-run results. Complete values are provided in Appendix E.6.

4.5 Ablation Study

RQ4 evaluates the deployment-time contribution of state conditioning and the three learned operators. All variants use the same Full checkpoint without retraining. Table 12 reports the numerical results, while Fig. 6 shows the mean-error change relative to Full on a logarithmic scale.



(a) MNIST neuroevolution

(b) Planar mechanical-arm optimization

Figure 5: Task-specific training beyond synthetic functions. (a) MNIST test accuracy under different fractions of training data. (b) Mean endpoint error for the simple and complex planar-arm cases; the horizontal axis is logarithmic. Markers and error bars report mean and standard deviation over five runs for GIFT-Evo, B2Opt, and the strongest remaining baseline. Small gray points show the means of the other baselines. The results are descriptive.

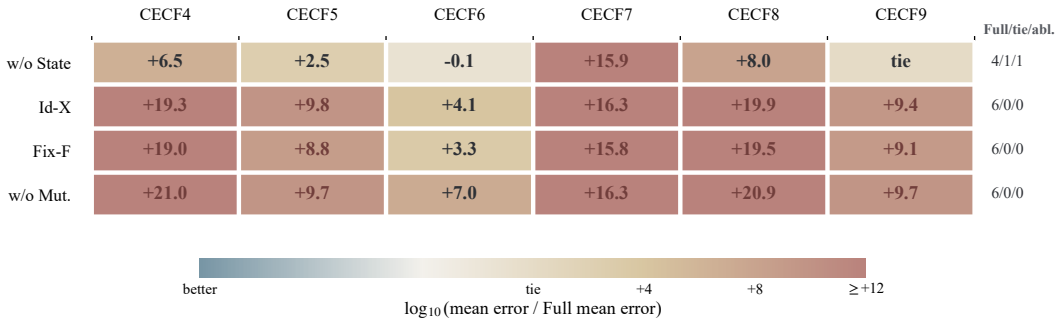


Figure 6: Deployment-time checkpoint ablation on CECF4–CECF9. Each cell reports the log-scale change in mean error relative to Full. Positive values indicate degradation. The final column gives Full-better/tie/ablation-better counts.

Removing state conditioning increases the error on four functions, improves CECF6, and ties on CECF9. Its benefit is therefore broad but function-dependent. Identity crossover, fixed fusion, and removal of mutation degrade performance on all six functions, often by several orders of magnitude. These results support the contributions of learned population mixing, adaptive fusion, mutation, and aggregate state conditioning. Complete results and definitions for all studies are provided in Appendix E.7.

5 Conclusion

We presented GIFT-Evo, a trainable evolutionary optimizer that combines geometric-inductive operator rules learned offline with search states accumulated during deployment. Its crossover, mutation, and fusion operators use relative population information, while individual acceptance, relative improvement, and rejection signals condition subsequent decisions without test-time parameter updates. Under controlled low-budget comparisons, GIFT-Evo achieves significant advantages in a majority of comparisons and directly deploys a frozen low-dimensional checkpoint at unseen higher dimensions; after task-specific training, it can also be instantiated for MNIST neuroevolution and planar-arm optimization. Future work will study more reliable state representations, scalable population mixing, and theoretical characterizations of the learned search dynamics.

AI USE STATEMENT

In this work, we used generative AI tools only to aid or polish the writing, including grammar correction, wording refinement, and improvements to readability. We did not use generative AI tools for any tasks requiring mandatory disclosure, including research ideation or execution, generating synthetic datasets, developing theoretical models or conceptual frameworks, formulating or proving mathematical claims, proposing or refining hypotheses, designing methodology or experiments, implementing methods, translating the manuscript, cleaning or reformatting datasets, conducting qualitative analysis, or interpreting results. We reviewed all AI-assisted revisions and verified that they did not alter the technical meaning, claims, results, or conclusions of the paper. We take responsibility for the final content of this work, including text produced with the aid of generative AI.

REFERENCES

Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies–a comprehensive introduction. *Natural computing*, 1(1):3–52, 2002.

Janez Brest, Sao Greiner, Borko Boskovic, Marjan Mernik, and Viljem Zumer. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE transactions on evolutionary computation*, 10(6):646–657, 2006.

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

Luis DaCosta, Alvaro Fialho, Marc Schoenauer, and Michèle Sebag. Adaptive operator selection with dynamic multi-armed bandits. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pp. 913–920, 2008.

Swagatam Das and Ponnuthurai Nagaratnam Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE transactions on evolutionary computation*, 15(1):4–31, 2010.

Steffen Finck, Nikolaus Hansen, Raymond Ros, and Anne Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical report, Citeseer, 2010.

Alexander IJ Forrester and Andy J Keane. Recent advances in surrogate-based optimization. *Progress in aerospace sciences*, 45(1-3):50–79, 2009.

Nikolaus Hansen. The cma evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Nikolaus Hansen and Andreas Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proceedings of IEEE international conference on evolutionary computation*, pp. 312–317. IEEE, 1996.

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

John H Holland. Outline for a logical theory of adaptive systems. *Journal of the ACM (JACM)*, 9 (3):297–314, 1962.

Frank Hutter, Lars Kotthoff, and Joaquin Vanschoren. *Automated machine learning: methods, systems, challenges*. Springer, 2019.

James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pp. 1942–1948. ieee, 1995.

Robert Lange, Tom Schaul, Yutian Chen, Chris Lu, Tom Zahavy, Valentin Dalibard, and Sebastian Flennerhag. Discovering attention-based genetic algorithms via meta-black-box optimization. In *Proceedings of the genetic and evolutionary computation conference*, pp. 929–937, 2023a.

Robert Lange, Tom Schaul, Yutian Chen, Tom Zahavy, Valentin Dalibard, Chris Lu, Satinder Singh, and Sebastian Flennerhag. Discovering evolution strategies via meta-black-box optimization. In *Proceedings of the companion conference on genetic and evolutionary computation*, pp. 29–30, 2023b.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Xiaobin Li, Kai Wu, Yujian B Li, Xiaoyu Zhang, Handing Wang, and Jing Liu. Pretrained optimization model for zero-shot black box optimization. *Advances in Neural Information Processing Systems*, 37:14283–14324, 2024.

Xiaobin Li, Kai Wu, Xiaoyu Zhang, and Handing Wang. B2opt: Learning to optimize black-box optimization with little budget. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18502–18510, 2025.

Zeyuan Ma, Hongshu Guo, Yue-Jiao Gong, Jun Zhang, and Kay Chen Tan. Toward automated algorithm design: A survey and practical guide to meta-black-box-optimization. *IEEE Transactions on Evolutionary Computation*, 2025.

Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

Felipe Petroski Such, Vashisht Madhavan, Edoardo Conti, Joel Lehman, Kenneth O Stanley, and Jeff Clune. Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. *arXiv preprint arXiv:1712.06567*, 2017.

Ryoji Tanabe and Alex S Fukunaga. Improving the search performance of shade using linear population size reduction. In *2014 IEEE congress on evolutionary computation (CEC)*, pp. 1658–1665. IEEE, 2014.

Ye Tian, Xuhong Qi, Shangshang Yang, Cheng He, Kay Chen Tan, Yaochu Jin, and Xingyi Zhang. A universal framework for automatically generating single-and multi-objective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 2025.

Chao Wang, Licheng Jiao, Lingling Li, Jiaxuan Zhao, Guanchun Wang, Fang Liu, and Shuyuan Yang. Task-free adaptive meta black-box optimization. *arXiv preprint arXiv:2601.21475*, 2026a.

Chao Wang, Lingling Li, Licheng Jiao, Jiaxuan Zhao, Fang Liu, and Shuyuan Yang. Learning evolution via optimization knowledge adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2026b. doi: 10.1109/TPAMI.2026.3686919.

Qing Ye, Yanan Sun, Jixin Zhang, and Jiancheng Lv. A distributed framework for ea-based nas. *IEEE Transactions on Parallel and Distributed Systems*, 32(7):1753–1764, 2020.

A Detailed Operator Parameterization

This appendix clarifies the representations and operator mechanisms summarized in Section 3. It complements the main-text equations without repeating the full evolutionary rollout.

A.1 Shared Relative Representations

At each generation, individuals are ordered by ascending fitness; the same permutation is applied to the population, fitness values, and search states. For sorted position i, we use the normalized rank and within-population fitness

$$r_i^t = \frac{i-1}{N-1}, \qquad \widetilde{F}_i^t = \frac{F_i^t - \min_j F_j^t}{\max_j F_j^t - \min_j F_j^t + \epsilon}. \tag{10}$$

Thus, rank encodes fitness order, while normalized fitness removes population-wise offsets and positive rescaling of raw objective values.

We characterize population geometry by its center and mean radius,

$$\bar{x}_t = \frac{1}{N}\sum_{j=1}^{N} x_j^t, \qquad \bar{\rho}_t = \frac{1}{N}\sum_{j=1}^{N} \|x_j^t - \bar{x}_t\|_2, \tag{11}$$

and use the relative radius

$$\rho_i^t = \frac{\|x_i^t - \bar{x}_t\|_2}{\bar{\rho}_t + \epsilon}. \tag{12}$$

Mutation receives the bounded form $\tanh(\rho_i^t)$, whereas fusion uses ρ_i^t directly. These features describe an individual relative to its current population rather than through raw fitness values or absolute location alone.

A.2 State-Conditioned Rank-Pair Crossover

Crossover determines how each target individual mixes information from the population. Its score for target–source pair (i, j) combines a learned rank-pair rule, a bounded correction based on normalized fitness and search states, and a learned self-retention term:

$$\begin{aligned} e_{ij}^t = {} & g_\theta^{\text{rank}}\big(r_i^t, r_j^t, |r_i^t - r_j^t|, \mathbb{I}[i=j]\big) \\ & + \tanh(\alpha)\, g_\theta^{\text{state}}\Big(\widetilde{F}_i^t, \widetilde{F}_j^t, s_i^t, s_j^t\Big) + b_{\text{self}}\mathbb{I}[i=j]. \end{aligned} \tag{13}$$

The scores are normalized over source individuals to form the mixing matrix:

$$A_{ij}^t = \frac{\exp(e_{ij}^t)}{\sum_{k=1}^{N}\exp(e_{ik}^t)}. \tag{14}$$

The regularizer in Eq. 9 penalizes the energy of the state-conditioned correction. It limits the correction magnitude without prescribing how a particular state should change a mixing weight.

For fixed fitness values and search states, a common translation $x_j^t \mapsto x_j^t + b$ leaves A_t unchanged. Therefore, the crossover candidate in Eq. 4 satisfies

$$\sum_{j=1}^{N} A_{ij}^t (x_j^t + b) = x_{\text{cross},i}^t + b. \tag{15}$$

This is a local translation-equivariance property of the crossover operation. It does not imply invariance or equivariance of mutation, fusion, boundary projection, or the complete optimization trajectory.

A.3 Geometry-Aware Mutation

Mutation complements crossover by introducing displacement beyond population mixing. It first forms a small set of learned population directions from centered individuals,

$$b_k^t = \sum_{j=1}^{N} \alpha_{kj}^t (x_j^t - \bar{x}_t), \qquad \sum_{j=1}^{N} \alpha_{kj}^t = 1, \tag{16}$$

where the weights depend on each individual's rank, normalized fitness, relative radius, and search state. These directions define a population basis B_t, from which the mutation module draws a learned exploratory displacement. A coordinate-wise component scaled by the current population spread is added to retain local exploration when the learned basis alone is insufficient.

Mutation also constructs a weighted reference m_t from the current population. A gated and bounded correction then points from the crossover candidate toward this reference:

$$u_i^t = \text{Cap}\big(g_i^t(m_t - x_{\text{cross},i}^t), \ell_i^t\big), \qquad \text{Cap}(u, \ell) = \min\left(1, \frac{\ell}{\|u\|_2 + \epsilon}\right) u. \tag{17}$$

The gate and bound are predicted from the same relative population information and search state. Consequently, Eq. 6 combines population-induced exploration with a bounded reference-guided correction, rather than using a fixed perturbation scale or a persistent solution archive.

A.4 Adaptive Fusion and Greedy Selection

Fusion receives normalized rank, search state, and relative population radius. As specified in Eq. 7, it adjusts a learnable base coefficient vector initialized to favor crossover. The three resulting coefficients weight the parent, crossover candidate, and mutation candidate. They are unconstrained linear coefficients: they need not be positive or sum to one, allowing preservation, interpolation, or extrapolation among the three branches.

The fused child is evaluated through the task-specific repair rule in Appendix E.2. Greedy parent–offspring selection retains the child whenever its evaluated fitness is no worse than that of its parent.

A.5 Feedback-Guided Search-State Update

Each individual starts with a zero search state. After evaluating child $x^t_{\mathrm{child},i}$, the feedback variables are

$$\begin{aligned} a_i^t &= \mathbb{I}\big[F^t_{\mathrm{child},i} \le F_i^t\big]\,, \\ \delta_i^t &= \frac{\max(0, F_i^t - F^t_{\mathrm{child},i})}{|F_i^t| + \epsilon}, \\ \widehat{s}_i^t &= \big[a_i^t, \tanh(\delta_i^t), 1 - a_i^t\big]\,, \\ s_i^{t+1} &= \mathrm{stopgrad}\big(\beta s_i^t + (1-\beta)\widehat{s}_i^t\big)\,. \end{aligned} \tag{18}$$

The three channels record acceptance, improvement magnitude, and rejection. The exponential moving average retains recent outcomes, while stop-gradient prevents backpropagation through the state sequence.

States remain attached to their surviving individuals when the population is reordered at the next generation. They condition crossover, mutation, and fusion in both offline rollouts and deployment; deployment updates only these states and keeps all learned operator parameters fixed.

B Task-Specific Training Objectives

Eq. 9 gives the common rollout-training objective. This section specifies the task losses used for synthetic functions, planar-arm optimization, and MNIST neuroevolution.

Synthetic proxy functions. For CECF1–CECF3, the task loss is the relative change in mean population fitness:

$$\mathcal{L}^{(f)}_{\mathrm{rel}} = \frac{\bar{f}(X_T) - \bar{f}(X_0)}{|\bar{f}(X_0)| + \epsilon}, \qquad \bar{f}(X_t) = \frac{1}{N}\sum_{i=1}^{N} f(x_i^t). \tag{19}$$

The regularized synthetic objective is

$$\mathcal{L}^{(f)}_{\mathrm{syn}} = \mathcal{L}^{(f)}_{\mathrm{rel}} + \lambda_{\mathrm{reg}}\mathcal{R}_{\mathrm{score}}, \tag{20}$$

where $\mathcal{R}_{\mathrm{score}}$ regularizes the state-conditioned crossover score. This objective rewards population improvement over an entire rollout rather than imitation of predefined evolutionary actions.

Planar-arm optimization. For target point $\tau = (\tau_x, \tau_y)$, the rollout fitness is endpoint error:

$$f_{\mathrm{arm}}(x;\tau) = \sqrt{(p_x(x) - \tau_x)^2 + (p_y(x) - \tau_y)^2 + 10^{-12}}, \tag{21}$$

where $p_x(x) = \sum_i L_i \cos(\theta_i)$ and $p_y(x) = \sum_i L_i \sin(\theta_i)$. The simple case optimizes joint angles with fixed $L_i = 10$; the complex case jointly optimizes $L_i \in [0, 10]$ and $\theta_i \in [-\pi, \pi]$.

A single GIFT-Evo model is jointly trained on both cases. Their relative-improvement losses and regularizers are averaged:

$$\mathcal{L}_{\mathrm{arm}} = \frac{1}{2}\sum_{c\in\{\mathrm{SC},\mathrm{CC}\}} \Big[\mathcal{L}^{(c)}_{\mathrm{rel}} + \lambda_{\mathrm{reg}}\mathcal{R}^{(c)}_{\mathrm{score}}\Big]\,. \tag{22}$$

Thus, endpoint error determines within-rollout ranking and selection, while its relative reduction trains the operator parameters.

Algorithm 1 One Generation of GIFT-Evo

Require: Population X_t, fitness F_t, states S_t, parameters θ, objective f
Ensure: Updated $(X_{t+1}, F_{t+1}, S_{t+1})$
1: Jointly sort (X_t, F_t, S_t) by ascending fitness
2: Construct rank, normalized-fitness, and relative-geometry features
3: Generate crossover and mutation candidates from (X_t, F_t, S_t) using θ
4: Fuse the parent, crossover, and mutation branches into X^t_{child}
5: Apply task-specific repair and evaluate F^t_{child}
6: Apply greedy parent–offspring selection to obtain (X_{t+1}, F_{t+1})
7: Update S_{t+1} from acceptance, relative improvement, and rejection outcomes
8: **return** $(X_{t+1}, F_{t+1}, S_{t+1})$

MNIST neuroevolution. Within a rollout, a candidate network with parameters w receives classification cross-entropy as fitness:

$$f_{\text{CE}}(w; \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(x,y)\in\mathcal{D}} \text{CE}(h_w(x), y)\,. \tag{23}$$

The outer objective combines final best fitness, mean child fitness before selection, final-population fitness, and regularization:

$$\begin{aligned}\mathcal{L}_{\text{MNIST}} = \mathbb{E}\Big[\min_i f_{\text{CE}}(w_i^T)\Big] + \lambda_{\text{cand}}\mathcal{L}_{\text{cand}} \\ + \lambda_{\text{pop}}\frac{1}{N}\sum_{i=1}^{N} f_{\text{CE}}(w_i^T) + \lambda_{\text{state}}\mathcal{R}_{\text{score}} + \lambda_{\text{mut}}\mathcal{R}_{\text{mut}},\end{aligned} \tag{24}$$

where

$$\mathcal{L}_{\text{cand}} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N} f_{\text{CE}}(w^t_{\text{child},i}) \tag{25}$$

is the rollout-average child-population fitness before greedy selection. The configured weights are reported in Appendix E. Classification accuracy is used only as the reported test metric, whose evaluation objective is $1 - \text{accuracy}$.

MNIST additionally uses a task-specific coordinate-aware candidate adapter after fusion and before fitness evaluation:

$$W^t_{\text{child,adapt}} = \Pi_{[-1,1]^D}\left[\text{CoordinateAdapter}\left(W^t_{\text{child}}; W_t, W^t_{\text{cross}}, W^t_{\text{mut}}, F_t, S_t\right)\right]. \tag{26}$$

The adapter is trained jointly with the evolutionary operators and remains fixed at deployment. It is specific to MNIST and is not part of the coordinate-free core operator design. A separate model is trained for each MNIST data fraction.

C Training and Deployment Algorithms

Training and deployment use the same generation-level transition. Algorithm 1 describes the shared operation; Algorithms 2 and 3 differ only in whether the completed rollout updates the learned parameters.

D Computational Complexity

For population size N, dimension D, and K learned mutation directions, the dominant operator cost is crossover mixing, $A_t X_t$, which requires $\mathcal{O}(N^2 D)$ time. Mutation contributes $\mathcal{O}(KND)$, while sorting and the remaining population statistics contribute $\mathcal{O}(N \log N + ND)$. Excluding objective-function evaluations, the per-generation overhead is

$$\mathcal{O}\big(N^2 D + KND + N\log N\big)\,. \tag{27}$$

For fixed K, crossover is the dominant term. This quadratic dependence on population size is the main computational limitation of the current population-mixing design; external objective evaluations remain separate because their cost is task-dependent.

Algorithm 2 Offline Training of GIFT-Evo

Require: Proxy-task set $\mathcal{F}_{\mathrm{tr}}$, rollout length T, population size N
Ensure: Trained operator parameters θ
Initialize θ
for each training iteration **do**
  Sample proxy tasks and initialize their populations
  Evaluate the initial populations and set $S_0 = \mathbf{0}$
  for $t = 0, \ldots, T-1$ **do**
    Apply Algorithm 1 to each rollout
  end for
  Compute the task-specific objective and regularization
  Update θ by gradient-based optimization
end for
return θ

Algorithm 3 Fixed-Parameter Deployment of GIFT-Evo

Require: Target objective f, trained parameters θ, budget B, population size N
Ensure: Best solution found
Load and fix θ; initialize X_0, evaluate F_0, and set $S_0 = \mathbf{0}$
while at least N evaluations remain **do**
  Apply Algorithm 1
end while
return the best individual in the final population

E Additional Experimental Details

This section summarizes the training configurations, benchmark protocols, and implementation details used in the reported experiments. Unless otherwise stated, the learned parameters remain fixed during deployment, and no test-function information is used for parameter updating or checkpoint selection.

E.1 Hyperparameter Configuration

The common operator and optimization settings are summarized in Table 3. Task-specific rollout configurations are reported separately in Table 4.

All task families are trained using Adam with the default PyTorch coefficients $(\beta_1, \beta_2) = (0.9, 0.999)$, numerical constant 10^{-8}, zero weight decay, and AMSGrad disabled. The learning rate is manually multiplied by 0.9 every 100 offline epochs, which is equivalent to an epoch-level StepLR schedule. For Synthetic, the schedule is applied once per epoch rather than after each task-wise or joint parameter update.

The number of offline epochs is therefore 500 for Synthetic and 1000 for Planar Arm. Each of the four MNIST data fractions is represented by an independently trained model, and each model is optimized for 1000 epochs. The offline epoch count E is distinct from the rollout length T.

Every offline rollout begins with

$$S_0 = \mathbf{0} \in \mathbb{R}^{M \times N \times 3}. \tag{28}$$

The state is propagated only across the optimizer steps of the current rollout. It is not carried across epochs, synthetic training functions, Planar Arm cases, or MNIST batches.

During offline training, the state update is detached before being passed to the next optimizer step. This prevents backpropagation through the recurrent state sequence while retaining the state as within-rollout search knowledge. During deployment, the same EMA update is used without gradient computation.

The mutation module uses $K_b = 4$ population-induced basis directions. Basis-coefficient perturbations and coordinate-wise perturbations are sampled during offline training. During deployment, the

Table 2: Trainable parameter counts for the default coordinate-free GIFT-Evo checkpoint used in the $d = 10$ controlled experiments.

Component	Parameters
State-conditioned rank-pair crossover	2,810
Geometry-aware mutation	13,818
Adaptive fusion	2,103
Total	18,731

Table 3: Common hyperparameter configuration of GIFT-Evo.

Hyperparameter	**Value**
Hidden dimension H	200
State dimension	3
State EMA coefficient β	0.8
Rank exponent	1.0
Population basis directions K_b	4
State regularization coefficient λ_s	10^{-4}
Initial fusion base coefficients	$[0, 1, 0]$
Optimizer	Adam
Initial learning rate	10^{-3}
Learning-rate decay	$\times 0.9$ every 100 epochs
Adam coefficients	$(0.9, 0.999)$
Adam numerical constant	10^{-8}
Weight decay	0
Global gradient-norm threshold	10
Mutation during training	Stochastic sampling
Mutation during deployment	Task-dependent; see protocols

means of the learned perturbation distributions are used, producing deterministic operator outputs for a fixed population, state, and objective function.

Synthetic and Planar Arm reuse one learned step module throughout their 30-step rollouts. MNIST uses three independently parameterized step modules, each with its own coordinate-aware adapter. The initialization probabilities of the adapter prior gates are listed in Table 5. These probabilities initialize learnable logits and are subsequently optimized during offline training.

E.2 Initial Populations and Search Domains

All initial populations are sampled independently across coordinates from task-dependent uniform distributions. The search domains are summarized in Table 6.

Initial fitness is calculated through the task evaluation map

$$F_0 = \mathcal{E}_f(X_0) = f\big(\mathrm{Repair}_f(X_0)\big)\,. \tag{29}$$

The repair operation used in this expression determines the coordinates on which fitness is evaluated but does not necessarily overwrite the population tensor.

For Synthetic and Planar Arm, boundary repair is applied internally during fitness evaluation. The optimizer retains the generated candidate tensor together with the fitness calculated from its repaired representation. For MNIST, the coordinate-aware adapter output is explicitly projected onto $[-1, 1]^{567}$ before evaluation and greedy selection.

E.3 Controlled BBOB Protocol

RQ1 and RQ2 use the 24-function BBOB implementation bundled with B2Opt. It follows the BBOB function family Finck et al. (2010), but is not an official COCO benchmark evaluation. All methods are compared at a reporting horizon of 3,100 function evaluations (FEs). The initial population

Table 4: Task-specific offline training configurations. Here, E denotes the number of offline epochs, M is the number of vectorized optimization instances, N is the population size, T is the rollout length, and B is the evaluation budget of one rollout.

Task	E	M	N	T	B	**Decision dimension**	**Step sharing**
Synthetic	500	64	100	30	3100	$D = 10$	Shared
Planar Arm	1000	64	100	30	3100	SC: 100; CC: 200	Shared
MNIST, each fraction	1000	1	36	3	144	$D = 567$	Not shared

Table 5: Initialization of the learnable MNIST coordinate-prior gates.

Training-data fraction	α_{init}
0.25	0.400
0.50	0.410
0.75	0.435
1.00	0.500

contains $N = 100$ candidates, and terminal errors and convergence curves are recorded at this horizon. At $d = 30$ and $d = 100$, native-configuration baseline trajectories are truncated at 3,100 FEs; no observation beyond this horizon enters the reported results or tests.

Each function–dimension–method setting uses 30 independent runs with seeds $1, \ldots, 30$, the same function implementation, instance, initialization domain $[-5, 5]^d$, and population size. BBOB evaluations and terminal-error summaries use double precision. GIFT-Evo uses one model trained on CECF1–CECF3 at $d = 10$, which remains frozen when directly deployed at $d = 10$, $d = 30$, and $d = 100$. B2Opt uses a separately trained checkpoint at each tested dimension and is therefore a performance baseline rather than a zero-shot cross-dimensional-transfer comparator.

For each baseline separately, we compare terminal errors on each of the 24 functions using two-sided Mann–Whitney U tests. The 24 p-values are Holm-corrected at $\alpha = 0.05$. We report $\mathrm{W}/\mathrm{NS}/\mathrm{L}$: a Holm-significant win, no significant difference, or a Holm-significant loss for GIFT-Evo. Complete terminal results, function-level outcomes, and convergence curves are provided in Appendix F.

ES uses a Geatpy-compatible $(\mu + \lambda)$-ES profile with $\lambda/N = 0.5$; DE uses DE/rand/1/bin with $F = 0.5$ and $\mathrm{CR} = 0.9$; PSO uses the global-best configuration; CMA-ES uses a uniformly sampled initial mean and step size $0.4(u - \ell)$; and L-SHADE uses the `pyade` profile with memory size 6. Table 7 summarizes the deployment treatment of learned baselines.

E.4 Planar Arm Protocol

The planar-arm task minimizes the Euclidean distance between the end effector and a two-dimensional target. We consider a simple case (SC), which optimizes 100 joint angles with all arm lengths fixed to 10, and a complex case (CC), which jointly optimizes 100 arm lengths and 100 joint angles. The corresponding decision dimensions are 100 and 200, respectively.

GIFT-Evo is trained from scratch for this task using a single checkpoint shared between SC and CC. The two cases use separate rollouts with zero initial search states, and their training losses are averaged before each parameter update. This experiment therefore evaluates task-specific adaptability rather than zero-shot transfer from the synthetic-function checkpoint.

During evaluation, targets are sampled within disks of radii $r \leq 100$, $r \leq 300$, and $r \leq 1000$. Each optimization run uses a population of 100 and a budget of 3,100 FEs. For each case–radius setting, we report the mean endpoint error and standard deviation over five independent runs. These results are descriptive and are not used for statistical superiority claims.

Table 6: Initial-population distributions used in training and evaluation.

Task	Initial-population distribution	Remarks
Synthetic training	$\mathcal{U}([-10, 10]^D)$	Applied to CECF1–CECF3
CECF4–CECF6	$\mathcal{U}([-100, 100]^D)$	Used in the ablation evaluation
CECF7	$\mathcal{U}([-5, 5]^D)$	Used in the ablation evaluation
CECF8	$\mathcal{U}([-600, 600]^D)$	Used in the ablation evaluation
CECF9	$\mathcal{U}([-32, 32]^D)$	Used in the ablation evaluation
BBOB	$\mathcal{U}([-5, 5]^D)$	Used for all 24 BBOB functions
Planar Arm SC	$\mathcal{U}([-\pi, \pi]^{100})$	The 100 arm lengths are fixed to 10
Planar Arm CC	$\mathcal{U}([0, 10]^{100}) \times \mathcal{U}([-\pi, \pi]^{100})$	Lengths and angles are jointly optimized
MNIST	$\mathcal{U}([-1, 1]^{567})$	Used during both training and testing

Table 7: Test-time treatment of learned BBOB baselines.

Method	Test-time treatment
LES	Fixed pretrained parameters; no parameter update.
LGA	Fixed pretrained parameters; no parameter update.
ABOM	Online parameter adaptation from target-search feedback.
B2Opt	Fixed pretrained parameters, using a separately trained model at each dimension.

E.5 MNIST Protocol

We optimize the 567 parameters of a CNN on MNIST. Four fractions of a fixed 1,000-example training pool are considered:

$$p_{\text{data}} \in \{0.25, 0.50, 0.75, 1.00\}.$$

A separate GIFT-Evo model is trained from scratch for each fraction. Each model contains three non-shared optimizer steps and the corresponding coordinate-aware adapters.

The offline MNIST rollouts minimize classification cross-entropy on minibatches drawn from the relevant training fraction. At test time, optimization uses training accuracy as the candidate-selection criterion. The selected candidate is then evaluated on a fixed 1,000-example test set; test accuracy is the reported metric. Thus, test labels are not used to guide the search or select a candidate.

Each test run uses a population of 36, three optimizer steps, and 144 function evaluations. Stochastic operator sampling remains enabled during deployment. For each data fraction, we report mean test accuracy and standard deviation over five independent runs with one restart per run. This experiment evaluates task-specific adaptability, and its five-run results are interpreted descriptively rather than as statistical superiority claims.

E.6 Complete Task-Specific Results

Tables 8 and 9 provide the values summarized in Fig. 5. All entries are descriptive mean (standard deviation) results over five runs.

E.7 Ablation Protocol and Complete Results

We conduct three ablation studies. The first two studies use the same GIFT-Evo checkpoint trained on CECF1–CECF3 at $d = 10$. We change only the deployment-time computation, and no variant is retrained.

The first study examines the contribution of each major component. w/o State sets the state inputs of all three operators to zero. Id-X replaces learned crossover with the parent population. Fix-F retains only the crossover proposal, and w/o Mut. removes the mutation branch.

The component results show that the full configuration performs well across the six tasks. Full obtains the best or tied-best mean on CECF4, CECF5, CECF7, CECF8, and CECF9. Removing

Table 8: MNIST test accuracy at each training-data fraction. Higher is better.

Method	0.25	0.50	0.75	1.00
ES	0.21 (0.02)	0.21 (0.01)	0.21 (0.01)	0.25 (0.02)
DE	0.21 (0.01)	0.24 (0.01)	0.22 (0.02)	0.23 (0.02)
PSO	0.16 (0.02)	0.17 (0.03)	0.19 (0.03)	0.20 (0.01)
CMA-ES	0.57 (0.04)	0.61 (0.03)	0.61 (0.03)	0.67 (0.01)
L-SHADE	0.31 (0.03)	0.29 (0.01)	0.30 (0.07)	0.30 (0.01)
LGA	0.53 (0.02)	0.51 (0.01)	0.52 (0.05)	0.54 (0.04)
LES	0.14 (0.02)	0.25 (0.03)	0.23 (0.07)	0.28 (0.01)
ABOM	0.15 (0.02)	0.19 (0.02)	0.20 (0.04)	0.19 (0.02)
B2Opt	0.69 (0.01)	**0.80 (0.01)**	**0.85 (0.00)**	0.86 (0.00)
GIFT-Evo	**0.73 (0.01)**	0.79 (0.01)	**0.85 (0.00)**	**0.87 (0.00)**

Table 9: Planar-arm endpoint error for the simple case (SC) and complex case (CC). Lower is better.

	SC			CC		
Method	$r = 100$	$r = 300$	$r = 1000$	$r = 100$	$r = 300$	$r = 1000$
ES	10.7 (0.4)	35.0 (2.4)	153.0 (2.2)	9.22 (0.5)	38.3 (0.6)	235.0 (1.4)
DE	1.60 (0.1)	2.65 (0.1)	155.0 (1.5)	1.17 (0.1)	13.6 (0.5)	278.0 (2.0)
PSO	1.46 (0.1)	1.12 (0.1)	72.1 (7.1)	0.66 (0.0)	0.49 (0.0)	248.0 (11.2)
CMA-ES	1.25 (0.1)	1.84 (0.1)	154.0 (2.3)	0.78 (0.1)	5.22 (0.5)	239.0 (0.8)
L-SHADE	0.66 (0.1)	0.54 (0.1)	**0.40 (0.1)**	0.25 (0.1)	54.5 (5.3)	218.0 (0.9)
LGA	178.0 (1.7)	219.0 (1.6)	446.0 (1.4)	71.5 (0.3)	129.0 (0.4)	394.0 (1.1)
LES	0.48 (0.1)	0.59 (0.1)	37.8 (1.9)	0.29 (0.1)	0.35 (0.1)	168.0 (3.3)
ABOM	1.54 (0.0)	2.01 (0.1)	165.0 (8.3)	1.15 (0.1)	14.6 (2.4)	341.0 (8.9)
B2Opt	0.19 (0.1)	0.28 (0.1)	15.1 (0.3)	0.15 (0.0)	**0.17 (0.0)**	17.9 (0.1)
GIFT-Evo	**0.17 (0.0)**	**0.19 (0.0)**	8.10 (2.97)	**0.14 (0.0)**	**0.17 (0.0)**	**9.80 (3.26)**

mutation causes a large error increase on CECF4, CECF6, and CECF8. Replacing learned crossover with the parent population also causes a clear performance drop. These results show that both learned crossover and mutation contribute to the final performance.

The second study examines how search state should be assigned to the three operators. w/o all state removes state inputs from all operators. w/o state to crossover, w/o state to mutation, and w/o state to fusion remove state inputs from the corresponding operator only. state only to crossover, state only to mutation, and state only to fusion provide state information only to the corresponding operator.

The state-assignment results show that the best state placement depends on the task. State only to fusion gives the lowest mean on CECF4. Full gives the lowest mean on CECF5. State only to crossover gives the lowest mean on CECF6. Several variants obtain zero on CECF7, while all variants have the same reported mean on CECF9. These results show that search state is useful, but its effect depends on the operator that receives it. Full provides the most stable performance across the six tasks.

The third study examines the contribution of relative population geometry to cross-dimensional deployment. Relative-Geometry Full uses the complete relative geometric representation. w/o Relative Radial Geometry removes the runtime radial and population-scale features and replaces the dynamic refinement bound with a fixed source-dimension scale. The model is trained at $d = 10$ and evaluated at $d = 10$, $d = 30$, and $d = 100$. The two variants use the same functions, evaluation budget, run seeds, and initial populations.

For each dimension, we compare 30 paired runs on 24 BBOB functions using a two-sided paired Wilcoxon signed-rank test. Holm correction is applied across the 24 functions separately for each dimension. In Table 10, W means that Relative-Geometry Full has a significantly lower terminal error, L means that w/o Relative Radial Geometry has a significantly lower terminal error, and NS denotes a non-significant difference.

The relative-geometry results show that Full obtains more significant wins than losses at all three dimensions. The W/NS/L counts are 12/5/7 at $d = 10$, 13/5/6 at $d = 30$, and 15/3/6 at $d = 100$. Across all 72 function–dimension comparisons, Full obtains 40 wins, 13 non-significant differences, and 19 losses. These results support the contribution of relative population-scale geometry to fixed-parameter transfer. They do not imply that the full model is better on every function.

Table 10: Relative-geometry ablation under fixed-parameter deployment. W/NS/L counts compare Relative-Geometry Full with w/o Relative Radial Geometry across 24 BBOB functions. Holm correction is applied separately for each dimension.

Dimension	Full wins	NS	Full losses	W/NS/L
$d = 10$	12	5	7	12/5/7
$d = 30$	13	5	6	13/5/6
$d = 100$	15	3	6	15/3/6
Overall	40	13	19	40/13/19

Table 11: State-part ablation on CECF4–CECF9. Values are mean (standard deviation) over 30 runs; lower is better. Bold and underlined entries denote the best and second-best mean values for each task.

Variant	CECF4	CECF5	CECF6	CECF7	CECF8	CECF9
Full	2.54E-18 (4.53E-18)	**3.38E-09** (3.20E-09)	8.87E+00 (4.93E-02)	**0.00E+00** (0.00E+00)	5.19E-20 (8.73E-20)	**9.54E-07** (0.00E+00)
w/o all state	8.99E-12 (3.49E-12)	1.67E-06 (2.68E-07)	7.00E+00 (4.92E-01)	2.18E+01 (5.16E+00)	3.49E-12 (2.72E-12)	**9.54E-07** (0.00E+00)
w/o state to crossover	1.79E-18 (2.55E-18)	4.40E-09 (4.48E-09)	8.87E+00 (3.97E-02)	**0.00E+00** (0.00E+00)	**3.76E-20** (7.02E-20)	**9.54E-07** (0.00E+00)
w/o state to mutation	1.76E-18 (2.79E-18)	4.20E-09 (3.92E-09)	8.86E+00 (4.90E-02)	**0.00E+00** (0.00E+00)	5.93E-20 (1.13E-19)	**9.54E-07** (0.00E+00)
w/o state to fusion	8.84E-12 (2.75E-12)	1.50E-06 (2.03E-07)	6.99E+00 (3.58E-01)	1.19E+01 (4.83E+00)	1.62E-12 (8.44E-13)	**9.54E-07** (0.00E+00)
state only to crossover	8.95E-12 (3.03E-12)	1.67E-06 (3.09E-07)	**6.87E+00** (8.01E-01)	2.12E+01 (4.43E+00)	4.93E-12 (3.29E-12)	**9.54E-07** (0.00E+00)
state only to mutation	9.62E-12 (2.71E-12)	1.42E-06 (2.91E-07)	7.00E+00 (4.95E-01)	1.25E+01 (6.37E+00)	1.21E-12 (5.82E-13)	**9.54E-07** (0.00E+00)
state only to fusion	**1.22E-18** (1.61E-18)	3.44E-09 (2.82E-09)	8.88E+00 (5.78E-02)	**0.00E+00** (0.00E+00)	3.96E-20 (6.76E-20)	**9.54E-07** (0.00E+00)

Each setting in the first two studies uses a population of 100, a 3,100-FE budget, and 30 independent runs. Table 12 reports the component ablation results. Table 11 reports the state-assignment results. Table 10 reports the relative-geometry results.

E.8 Implementation and Reproducibility

GIFT-Evo is implemented in PyTorch. The training and evaluation scripts record the checkpoint, task configuration, seed protocol, population size, evaluation budget, and tested dimension for each experiment. For the controlled BBOB experiments, runs use seeds $1, \ldots, 30$; the task-specific experiments use five independent runs.

GIFT-Evo parameters are optimized only during offline training and remain fixed during deployment. Each rollout starts from a freshly sampled initial population and $S_0 = \mathbf{0}$. A parameter update is accepted only when both the training loss and global gradient norm are finite; gradients are clipped to a global ℓ_2-norm of 10.

For GIFT-Evo, evaluating the initial population costs N FEs. Each executed generation forms one fused offspring per parent, applies the task-specific repair rule for evaluation and greedy selection, and costs a further N FEs. Thus, the reported GIFT-Evo configurations satisfy $B = N(1 + T)$. Population coordinates, fitness values, and search states are jointly reordered by fitness before every generation, preserving the alignment between each population position and its state. Greedy selection ensures that the best terminal-population fitness equals the best-so-far fitness over the executed rollout.

Table 12: Deployment-time checkpoint ablation on CECF4–CECF9. Values are the mean (standard deviation) over 30 runs; lower is better.

Model	CECF4	CECF5	CECF6	CECF7	CECF8	CECF9
Full	**2.71e-18** (7.69e-18)	**5.22e-09** (4.23e-09)	8.87e+00 (3.32e-02)	**0.00e+00** (0.00e+00)	**3.78e-20** (6.03e-20)	**9.54e-07** (0.00e+00)
w/o State	9.01e-12 (3.92e-12)	1.71e-06 (2.80e-07)	**6.98e+00** (4.89e-01)	2.27e+01 (4.87e+00)	3.54e-12 (3.08e-12)	**9.54e-07** (0.00e+00)
Id-X	4.98e+01 (1.70e+01)	3.00e+01 (1.54e+01)	1.07e+05 (1.58e+05)	5.69e+01 (1.68e+01)	3.19e+00 (7.86e+00)	8.06e+00 (7.73e-01)
Fix-F	2.92e+01 (7.41e+00)	3.01e+00 (6.69e-01)	1.60e+04 (8.09e+03)	1.83e+01 (5.37e+00)	1.19e+00 (7.32e-02)	3.56e+00 (3.82e-01)
w/o Mut.	2.82e+03 (5.86e+02)	2.79e+01 (3.65e+00)	9.63e+07 (5.09e+07)	5.30e+01 (6.39e+00)	2.76e+01 (6.16e+00)	1.45e+01 (9.49e-01)

Bold and underlined entries denote the best and second-best reported means. For CECF7, the Full mean rounds to zero at the reported precision; its positive double-precision value is used only for the logarithmic ratio in Figure 6. CECF9 is treated as a tie according to the reported means.

Table 13: Holm-corrected $\mathrm{W}/\mathrm{NS}/\mathrm{L}$ outcomes for GIFT-Evo against each baseline under the common 3,100-FE budget.

Baseline	$d = 10$	$d = 30$	$d = 100$
ES	24/0/0	24/0/0	22/0/2
DE	24/0/0	24/0/0	24/0/0
PSO	23/1/0	23/1/0	21/0/3
CMA-ES	24/0/0	24/0/0	23/1/0
L-SHADE	23/1/0	24/0/0	22/2/0
LES	22/2/0	21/2/1	21/0/3
LGA	23/1/0	24/0/0	22/2/0
ABOM	23/1/0	24/0/0	23/1/0
B2Opt	24/0/0	24/0/0	24/0/0

F Additional BBOB Results

F.1 Holm-Corrected Outcomes across Dimensions

Table 13 expands the aggregate RQ2 counts by baseline. Each entry reports significant wins, non-significant differences, and significant losses for GIFT-Evo over the 24 functions.

F.2 Complete Controlled-Budget Convergence Curves

Fig. 7, 8, and 9 show the median best-so-far error over 30 independent runs for every BBOB function at the 3,100-FE reporting horizon. The $d = 30$ and $d = 100$ panels provide the complete visual evidence for cross-dimensional deployment without duplicating terminal tables.

G Supplementary Native-Budget Scalability Results

This section is separated from the controlled 3,100-FE evaluation because the evaluation budgets differ across methods. GIFT-Evo and B2Opt use 3,100 FEs, whereas conventional baselines use their method-specific budget of $310d$ FEs where available. Bold and underlined values identify the two lowest raw means within each table only. They do not support a fair-budget ranking or a statistical superiority claim.

Native-budget results at $d = 30$ and $d = 100$. Table 14 reports the native-budget results at $d = 30$ and $d = 100$. The conventional baseline budget is 9,300 FEs at $d = 30$ and 31,000 FEs at $d = 100$, while GIFT-Evo and B2Opt use 3,100 FEs. It therefore characterizes behavior under asymmetric evaluation allowances.

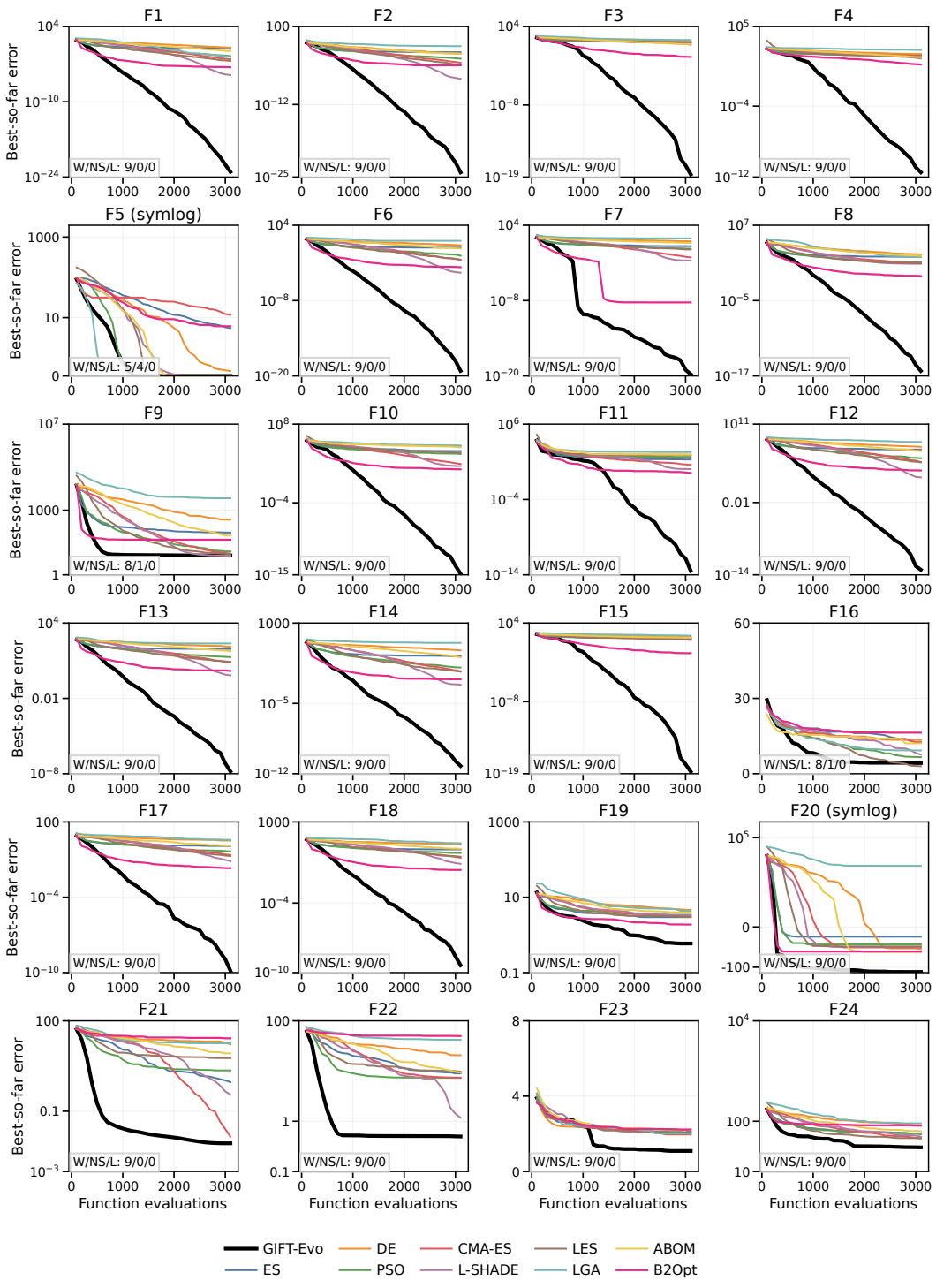


Figure 7: Median best-so-far error over 30 runs for all 24 BBOB functions at $d = 10$, reported at the 3,100-FE horizon.

Native-budget results at $d = 500$. Table 15 reports the native-budget results at $d = 500$. The conventional baseline budget is 155,000 FEs, while GIFT-Evo and B2Opt use 3,100 FEs. It remains a descriptive low-budget diagnostic under an asymmetric protocol.

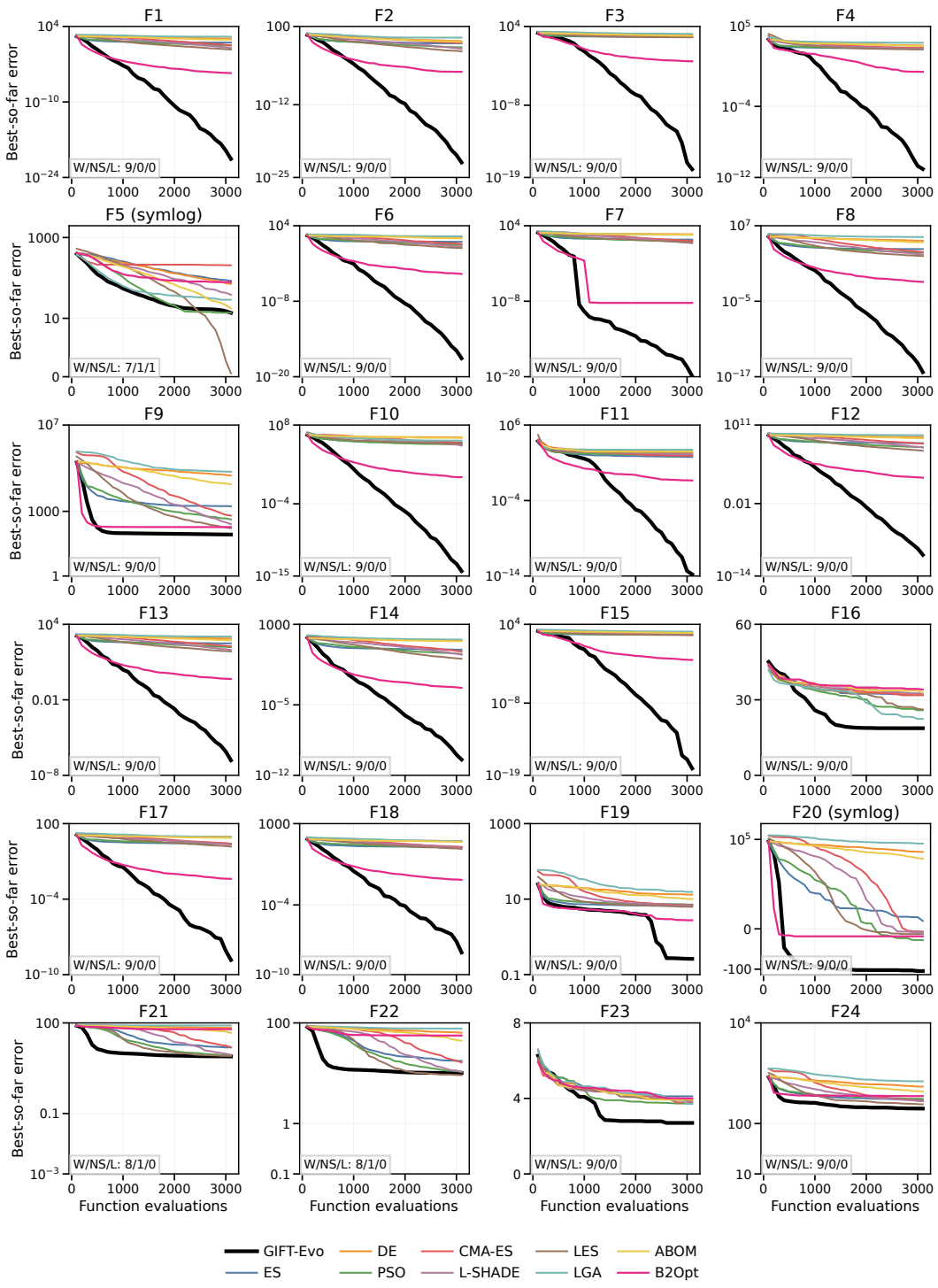


Figure 8: Median best-so-far error over 30 runs for all 24 BBOB functions at $d = 30$, reported at the 3,100-FE horizon.

Native-budget results at $d = 1000$. Table 16 reports the native-budget results at $d = 1000$. ES, DE, PSO, L-SHADE, LES, LGA, and ABOM use 310,000 FEs, whereas GIFT-Evo uses 3,100

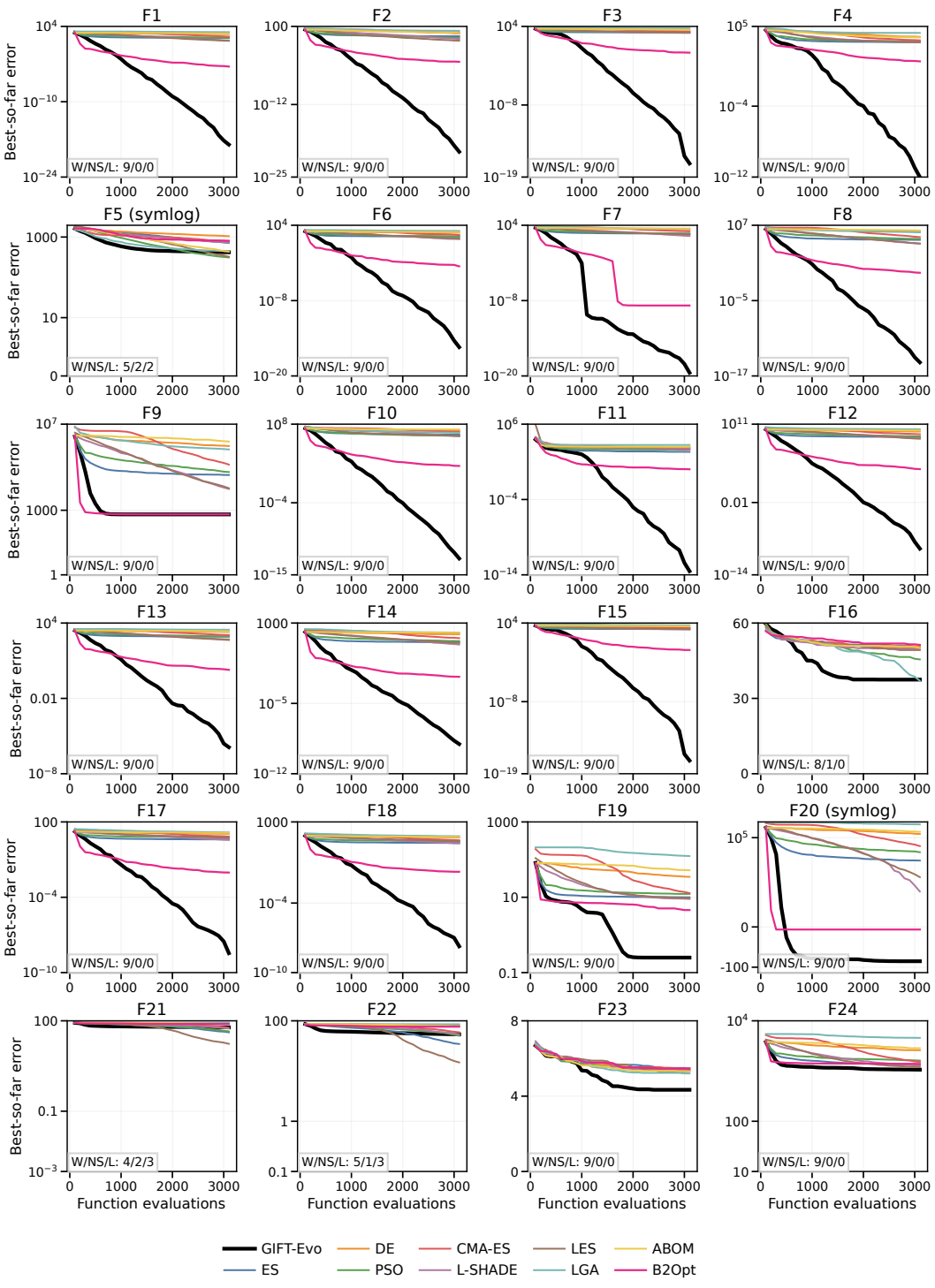


Figure 9: Median best-so-far error over 30 runs for all 24 BBOB functions at $d = 100$, reported at the 3,100-FE horizon.

FEs. CMA-ES and B2Opt could not be executed at this dimensionality; they remain visible as N/A columns and are excluded from every comparison.

Table 14: Native-budget BBOB results at $d = 30$ and $d = 100$. Entries are mean (standard deviation); lower is better. GIFT-Evo and B2Opt use 3,100 FEs, while conventional baselines use 9,300 FEs at $d = 30$ and 31,000 FEs at $d = 100$. Bold and underlined values mark the two lowest raw means under this asymmetric protocol. The final row for each dimension gives Holm-corrected W/NS/L counts for GIFT-Evo against each baseline across 24 functions.

F	GIFT-Evo	ES	DE	PSO	CMA-ES	L-SHADE	LES	LGA	ABOM	B2Opt
					Dimension $D = 30$					
F01	**4.065E-20 (1.237E-19)**	8.455E+00 (1.037E+00)	9.972E+00 (2.514E+00)	6.618E-02 (4.771E-02)	3.425E-06 (1.844E-06)	1.832E-01 (9.461E-02)	3.050E+00 (6.342E+00)	1.179E+02 (4.948E+01)	9.341E-09 (7.645E-10)	6.090E-06 (1.700E-06)
F02	**1.664E-22 (2.830E-22)**	7.554E-02 (1.505E-02)	1.264E-02 (2.322E-03)	4.853E-04 (3.633E-04)	5.550E-03 (2.784E-03)	7.047E-04 (3.658E-04)	5.372E-02 (1.101E-01)	1.063E+00 (6.069E-01)	3.903E-07 (5.870E-07)	1.700E-07 (3.400E-08)
F03	**1.009E-17 (2.239E-17)**	2.427E+02 (1.509E+01)	2.898E+02 (1.920E+01)	1.294E+02 (3.716E+01)	7.799E+01 (4.148E+01)	1.540E+02 (2.000E+01)	2.638E+02 (5.487E+01)	7.127E+02 (2.034E+02)	3.246E+01 (1.062E+01)	5.420E-03 (1.820E-03)
F04	**2.970E-11 (7.001E-11)**	3.060E+02 (1.739E+01)	3.958E+02 (2.565E+01)	1.568E+02 (2.743E+01)	9.185E+01 (3.448E+01)	2.131E+02 (2.426E+01)	5.405E+02 (7.507E+02)	1.432E+03 (5.999E+02)	3.940E+01 (1.195E+01)	7.370E-02 (2.880E-02)
F05	1.357E+01 (7.677E+00)	2.863E+01 (4.374E+00)	2.648E+01 (5.302E+00)	1.189E+01 (9.933E+00)	1.205E+01 (5.722E+00)	1.560E+00 (1.480E+00)	4.804E+01 (2.987E+01)	2.806E+01 (2.387E+01)	**0.000E+00 (0.000E+00)**	1.070E+01 (1.090E+01)
F06	**4.268E-17 (8.253E-17)**	2.020E+01 (2.846E+00)	3.209E+01 (6.882E+00)	5.417E-01 (2.583E-01)	3.021E-04 (1.732E-04)	1.409E+00 (5.935E-01)	1.720E+01 (1.930E+01)	2.271E+02 (7.915E+01)	1.058E-08 (3.475E-09)	8.820E-05 (2.500E-05)
F07	**1.736E-20 (2.057E-20)**	4.659E+01 (7.684E+00)	1.206E+02 (2.096E+01)	1.259E+01 (5.157E+00)	4.937E+00 (2.961E+00)	5.177E+00 (2.180E+00)	7.886E+01 (4.039E+01)	3.672E+02 (9.778E+01)	2.290E+01 (1.252E+01)	1.260E-12 (9.390E-13)
F08	**3.989E-16 (9.157E-16)**	1.207E+03 (2.406E+02)	2.473E+03 (9.280E+02)	9.533E+01 (4.409E+01)	2.931E+01 (5.259E+00)	1.237E+02 (2.670E+01)	2.229E+03 (6.041E+03)	1.638E+05 (8.289E+04)	7.946E+01 (6.627E+01)	3.890E-03 (9.110E-04)
F09	8.271E+01 (1.217E+01)	1.257E+03 (2.181E+02)	4.200E+03 (1.682E+03)	7.462E+01 (3.999E+01)	**3.645E+01 (2.150E+01)**	9.807E+01 (3.669E+01)	1.364E+03 (2.416E+03)	7.322E+04 (4.257E+04)	6.915E+01 (1.468E+02)	1.810E+02 (1.470E+00)
F10	**4.244E-14 (1.012E-13)**	1.749E+05 (3.368E+04)	3.253E+05 (6.793E+04)	4.088E+04 (2.219E+04)	8.999E+03 (4.233E+03)	1.054E+04 (5.672E+03)	1.474E+05 (8.082E+04)	3.807E+05 (1.910E+05)	3.596E+04 (1.368E+04)	5.920E-01 (1.360E-01)
F11	**4.062E-14 (6.566E-14)**	4.413E+01 (1.016E+01)	2.338E+02 (3.512E+01)	9.039E+01 (3.178E+01)	1.265E+02 (2.541E+01)	3.344E+01 (2.589E+01)	2.773E+02 (5.179E+01)	5.433E+02 (1.143E+02)	2.509E+02 (4.917E+01)	3.900E-02 (1.660E-02)
F12	**8.532E-11 (1.252E-10)**	7.733E+07 (7.150E+06)	1.618E+08 (4.199E+07)	4.413E+05 (2.524E+05)	8.277E+01 (6.245E+01)	3.289E+06 (2.050E+06)	6.487E+07 (1.049E+08)	1.859E+09 (1.170E+09)	1.618E+03 (6.569E+03)	6.210E+01 (2.320E+01)
F13	**2.781E-07 (3.206E-07)**	2.781E+02 (1.844E+01)	3.306E+02 (4.252E+01)	2.179E+01 (6.409E+00)	1.882E+00 (2.036E+00)	3.611E+01 (6.939E+00)	1.505E+02 (1.153E+02)	1.007E+03 (2.158E+02)	5.451E+00 (8.422E+00)	2.740E-01 (4.920E-02)
F14	**7.245E-11 (1.079E-10)**	2.540E+02 (2.821E+01)	7.071E+02 (1.443E+02)	6.788E-02 (2.944E-02)	6.698E-03 (3.049E-03)	2.157E-01 (1.307E-01)	5.109E+01 (4.652E+00)	2.966E+03 (6.866E+00)	2.856E-03 (5.170E-04)	3.340E-04 (4.460E-05)
F15	**6.885E-18 (1.253E-17)**	2.485E+02 (1.283E+01)	3.232E+02 (2.161E+01)	2.009E+02 (2.884E+01)	1.325E+02 (6.543E+01)	1.845E+02 (1.985E+01)	3.388E+02 (5.430E+01)	7.308E+02 (1.954E+02)	1.153E+02 (3.789E+01)	1.300E-02 (3.930E-03)
F16	1.767E+01 (2.563E+00)	2.917E+01 (2.748E+00)	2.979E+01 (2.299E+00)	1.200E+01 (5.264E+00)	**2.958E+00 (1.815E+00)**	2.522E+01 (3.899E+00)	1.402E+01 (2.933E+00)	1.974E+01 (8.278E+00)	1.262E+01 (9.104E+00)	2.450E+01 (1.120E+00)
F17	**2.271E-09 (2.442E-09)**	2.231E+00 (1.418E-01)	5.120E+00 (6.048E-01)	1.338E+00 (3.820E-01)	8.523E-02 (6.578E-02)	8.965E-01 (3.488E-01)	5.476E+00 (1.581E+00)	7.772E+00 (1.519E+00)	2.151E+00 (7.470E-01)	2.540E-03 (3.730E-04)
F18	**1.196E-08 (1.185E-08)**	8.628E+00 (5.573E-01)	1.795E+01 (2.300E+00)	5.700E+00 (2.405E+00)	7.149E-01 (3.578E-01)	3.050E+00 (5.961E-01)	1.968E+01 (5.948E+00)	2.587E+01 (7.511E+00)	7.264E+00 (2.643E+00)	7.410E-03 (1.450E-03)
F19	**5.673E-01 (7.711E-01)**	6.076E+00 (4.094E-01)	8.041E+00 (6.100E-01)	6.370E+00 (4.882E-01)	6.317E+00 (4.701E-01)	6.345E+00 (4.521E-01)	8.072E+00 (8.663E-01)	1.408E+01 (6.016E+00)	5.713E+00 (4.910E-01)	7.710E-01 (3.390E-01)
F20	**-1.262E+02 (5.142E+01)**	-9.547E+01 (1.301E+00)	8.085E+01 (1.552E+02)	-5.751E+00 (6.968E+00)	-8.217E+01 (1.150E+00)	-6.848E+01 (1.574E+00)	1.848E+03 (3.389E+03)	5.202E+04 (3.408E+04)	-1.523E+00 (1.697E+00)	7.630E+00 (3.080E+01)
F21	**6.726E+00 (2.059E+00)**	1.265E+01 (9.324E-01)	1.343E+01 (6.298E+00)	7.601E+00 (7.067E+00)	8.138E+00 (4.690E+00)	7.817E+00 (4.164E-01)	1.970E+01 (1.778E+01)	7.708E+01 (1.254E+01)	8.550E+00 (5.615E+00)	2.000E+01 (2.620E+00)
F22	9.356E+00 (1.868E+00)	1.540E+01 (6.589E-01)	1.749E+01 (6.539E+00)	**8.771E+00 (3.739E+00)**	1.031E+01 (4.631E+00)	9.139E+00 (2.319E-01)	2.388E+01 (1.436E+01)	7.383E+01 (1.277E+01)	9.002E+00 (3.868E+00)	5.020E+01 (2.820E+00)
F23	2.646E+00 (2.849E-01)	3.406E+00 (4.254E-01)	3.330E+00 (4.437E-01)	3.234E+00 (4.698E-01)	4.037E+00 (5.145E-01)	4.036E+00 (5.218E-01)	4.646E+00 (6.199E-01)	2.830E+00 (1.345E+00)	3.268E+00 (4.909E-01)	**2.370E+00 (4.190E-01)**
F24	**1.973E+02 (1.022E+01)**	2.732E+02 (1.195E+01)	3.347E+02 (1.920E+01)	2.687E+02 (3.097E+01)	2.538E+02 (1.630E+01)	2.500E+02 (1.662E+01)	3.629E+02 (4.839E+01)	5.674E+02 (1.903E+02)	2.121E+02 (5.260E+01)	3.080E+02 (8.780E+00)
GIFT-Evo W/NS/L	–	24/0/0	23/0/1	19/3/2	21/1/2	18/2/4	18/3/3	21/3/0	22/1/1	22/2/0
F	GIFT-Evo	ES	DE	PSO	CMA-ES	L-SHADE	LES	LGA	ABOM	B2Opt
					Dimension $D = 100$					
F01	**2.321E-18 (3.328E-18)**	5.524E+01 (2.921E+00)	1.605E+01 (2.207E+00)	1.220E+00 (6.845E-01)	1.231E-07 (5.254E-08)	6.690E-02 (3.621E-02)	3.505E-08 (1.786E-09)	7.683E+02 (1.233E+02)	1.170E-01 (2.176E-01)	8.530E-05 (1.690E-05)
F02	**2.564E-20 (6.463E-20)**	1.194E+00 (1.038E-01)	4.743E-02 (9.641E-03)	1.203E-02 (1.721E-02)	1.520E-02 (4.603E-03)	6.363E-04 (5.893E-04)	2.486E-03 (1.554E-02)	1.825E+01 (1.051E+01)	5.937E-02 (1.817E-02)	7.180E-06 (2.020E-06)
F03	**4.626E-17 (1.114E-16)**	1.076E+03 (3.297E+01)	1.028E+03 (3.666E+01)	4.988E+02 (7.586E+01)	8.339E+01 (3.060E+01)	3.039E+02 (2.925E+01)	1.447E+02 (3.480E+01)	5.058E+03 (6.584E+02)	2.770E+02 (3.517E+01)	2.480E-01 (5.890E-02)
F04	**3.499E-12 (8.017E-12)**	1.412E+03 (4.363E+01)	1.462E+03 (5.549E+01)	8.840E+02 (1.826E+02)	1.308E+02 (1.598E+01)	4.281E+02 (5.053E+01)	2.415E+02 (8.680E+01)	1.831E+04 (8.328E+03)	6.083E+02 (8.989E+01)	2.440E+00 (9.820E-01)
F05	4.120E+02 (5.230E+01)	1.890E+02 (9.929E+00)	1.243E+02 (1.688E+01)	1.992E+02 (9.061E+01)	2.045E+02 (3.766E+01)	**4.950E-01 (2.711E+00)**	2.325E+00 (7.223E+00)	3.952E+02 (1.407E+02)	1.269E+00 (5.886E-01)	5.360E+02 (5.180E+01)
F06	**1.255E-15 (1.946E-15)**	1.167E+02 (5.865E+00)	4.775E+01 (9.854E+00)	8.732E+00 (4.075E+00)	5.006E-05 (2.248E-05)	1.459E+00 (6.734E-01)	1.300E-06 (2.727E-07)	1.111E+03 (1.633E+02)	4.343E+00 (2.683E+00)	8.010E-04 (3.320E-04)
F07	**3.910E-20 (3.673E-20)**	3.702E+02 (1.845E+01)	2.833E+02 (5.134E+01)	1.322E+02 (3.127E+01)	5.522E+00 (1.531E+00)	2.914E+01 (7.599E+00)	4.065E+01 (1.939E+01)	2.587E+03 (6.359E+02)	2.561E+02 (6.791E+01)	1.210E-12 (1.410E-12)
F08	**6.126E-15 (9.273E-15)**	2.820E+04 (2.465E+03)	9.385E+03 (2.557E+03)	9.420E+02 (2.108E+02)	9.679E+01 (8.242E-01)	4.654E+02 (6.762E+01)	1.284E+02 (5.053E+01)	8.734E+05 (1.109E+06)	4.015E+03 (1.650E+03)	7.000E-02 (2.220E-02)
F09	6.419E+02 (4.860E-01)	2.709E+04 (2.724E+03)	6.789E+03 (2.182E+03)	1.036E+03 (7.447E+02)	**9.656E+01 (5.266E-01)**	2.968E+02 (8.096E+01)	1.178E+02 (6.836E+01)	7.647E+05 (1.060E+06)	4.566E+02 (1.901E+02)	6.420E+02 (2.860E-01)
F10	**1.351E-12 (2.079E-12)**	1.769E+06 (2.204E+05)	4.030E+06 (5.186E+05)	2.130E+05 (5.228E+04)	1.887E+04 (5.505E+03)	2.717E+04 (1.005E+04)	1.182E+05 (3.626E+04)	3.659E+06 (2.597E+06)	2.440E+06 (5.133E+05)	7.000E+00 (1.850E+00)
F11	**1.100E-13 (3.129E-13)**	1.502E+02 (1.069E+01)	8.058E+02 (9.087E+01)	3.041E+02 (7.039E+01)	2.339E+02 (2.292E+01)	7.138E+01 (1.886E+02)	8.344E+02 (7.609E+01)	1.706E+03 (3.130E+02)	8.921E+02 (7.160E+01)	3.550E-02 (2.070E-02)
F12	**5.927E-10 (8.594E-10)**	5.672E+08 (3.352E+07)	3.225E+08 (6.881E+07)	1.267E+07 (7.245E+06)	4.673E+06 (2.421E+06)	1.224E+06 (8.125E+05)	3.644E+07 (4.938E+00)	1.277E+10 (5.696E+09)	5.467E+06 (3.945E+06)	9.970E+02 (2.370E+02)
F13	**1.542E-06 (1.221E-06)**	7.362E+02 (2.302E+01)	3.948E+02 (3.628E+01)	1.059E+02 (2.744E+01)	2.654E-01 (2.682E-01)	2.625E+01 (1.052E+01)	1.286E+00 (1.905E+00)	2.803E+03 (2.123E+02)	3.550E+01 (1.462E+01)	1.020E+00 (1.480E-01)
F14	**1.106E-09 (9.740E-10)**	9.349E+00 (5.422E-01)	9.402E+00 (1.524E+00)	6.333E-01 (2.439E-01)	8.881E-03 (1.144E-03)	1.027E-01 (3.061E-02)	1.237E-01 (6.435E-01)	1.019E+02 (2.113E+01)	2.426E+00 (2.076E+00)	1.620E-03 (2.970E-04)
F15	**5.548E-17 (9.903E-17)**	1.090E+03 (2.868E+01)	1.101E+03 (3.416E+01)	7.673E+02 (1.314E+02)	3.533E+02 (3.072E+02)	3.980E+02 (3.315E+01)	1.240E+02 (2.936E+01)	4.232E+03 (4.680E+02)	9.649E+02 (3.247E+01)	2.020E-01 (6.580E-02)
F16	3.725E+01 (3.482E+00)	4.569E+01 (2.154E+00)	4.622E+01 (1.781E+00)	1.923E+01 (5.010E+00)	8.315E+00 (1.095E+01)	3.272E+01 (5.127E+00)	**1.495E+00 (4.088E-01)**	3.092E+01 (1.021E+01)	4.623E+01 (1.558E+00)	4.440E+01 (1.170E+00)
F17	**4.466E-09 (4.753E-09)**	3.473E+00 (1.203E-01)	4.007E+00 (4.710E-01)	3.500E+00 (5.738E-01)	4.268E-03 (1.148E-03)	9.330E-01 (1.213E-01)	4.250E-01 (2.798E-01)	1.294E+01 (1.177E+00)	1.823E+00 (5.245E-01)	4.070E-03 (5.720E-04)
F18	**2.341E-08 (2.301E-08)**	1.325E+01 (7.007E-01)	1.592E+01 (1.754E+00)	1.274E+01 (2.751E+00)	6.304E-02 (1.572E-02)	3.356E+00 (7.329E-01)	1.857E+00 (8.920E-01)	4.977E+01 (6.495E+00)	1.148E+01 (2.188E+00)	1.950E-02 (2.130E-03)
F19	**2.502E-01 (8.614E-05)**	9.084E+00 (2.517E-01)	9.023E+00 (3.345E-01)	9.308E+00 (6.239E-01)	7.244E+00 (3.436E-01)	6.970E+00 (4.465E-01)	7.321E+00 (2.515E-01)	1.031E+02 (2.988E+01)	8.254E+00 (2.577E-01)	1.430E+00 (4.880E-01)
F20	**-5.108E+01 (2.334E+01)**	1.429E+03 (4.209E+02)	3.757E+01 (4.272E+01)	3.005E+02 (9.060E+02)	-6.505E-01 (6.147E-01)	-3.919E-01 (9.664E-01)	3.629E+02 (1.755E+03)	4.380E+05 (9.610E+04)	-1.596E+00 (1.438E+00)	4.390E-01 (3.470E+00)
F21	5.957E+01 (4.667E+00)	1.690E+01 (5.130E+00)	7.832E+00 (4.976E+00)	6.045E+00 (5.712E+00)	5.965E+00 (5.072E+00)	5.036E+00 (3.857E+00)	1.035E+01 (1.001E+01)	8.482E+01 (2.415E+00)	**4.055E+00 (4.050E+00)**	7.200E+01 (6.660E-01)
F22	5.414E+01 (2.689E+00)	1.829E+01 (6.924E-01)	1.036E+01 (6.841E-01)	9.482E+00 (8.120E+00)	**7.272E+00 (1.785E-08)**	7.420E+00 (1.036E-01)	1.461E+01 (1.682E+01)	8.455E+01 (2.906E+00)	8.778E+00 (6.803E-01)	7.780E+01 (1.590E+00)
F23	4.301E+00 (2.973E-01)	4.673E+00 (3.446E-01)	4.734E+00 (2.842E-01)	4.210E+00 (4.522E-01)	4.840E+00 (2.410E-01)	4.695E+00 (4.666E-01)	4.708E+00 (3.22E-01)	**3.812E+00 (1.568E+00)**	4.792E+00 (2.581E-01)	4.230E+00 (7.640E-02)
F24	1.067E+03 (3.064E+01)	1.180E+03 (3.580E+01)	1.176E+03 (4.488E+01)	1.292E+03 (9.938E+01)	8.866E+02 (2.751E+01)	**7.425E+02 (7.003E+01)**	8.834E+02 (5.810E+01)	4.042E+03 (6.390E+02)	1.073E+03 (2.479E+01)	1.300E+03 (3.370E+01)
GIFT-Evo W/NS/L	–	21/0/3	21/0/3	19/1/4	18/0/6	18/0/6	18/0/6	21/2/1	19/1/4	22/2/0

Table 15: Terminal BBOB error under method-specific native budgets at $d = 500$. Entries are mean (standard deviation); lower is better. GIFT-Evo and B2Opt use 3,100 FEs, whereas ES, DE, PSO, CMA-ES, L-SHADE, LES, LGA, and ABOM use 155,000 FEs. The final row gives Holm-corrected W/NS/L counts under this asymmetric-budget diagnostic; it is not a common-budget comparison.

F	GIFT-Evo	ES	DE	PSO	CMA-ES	L-SHADE	LES	LGA	ABOM	B2Opt
F01	**3.541E-17 (9.437E-17)**	4.588E+02 (9.346E+00)	6.791E+01 (4.420E+00)	1.482E+02 (2.972E+01)	2.480E+03 (8.030E+01)	1.055E+01 (1.755E+00)	6.502E+01 (4.541E+01)	6.185E+03 (1.425E+02)	1.055E+03 (2.595E+02)	9.810E-05 (1.600E-05)
F02	**5.412E-19 (6.241E-19)**	1.775E+01 (8.360E-01)	2.117E-01 (2.854E-02)	8.051E-01 (2.162E-01)	2.110E+02 (1.230E+01)	9.208E-02 (3.517E-02)	6.971E-01 (4.034E-01)	2.830E+02 (2.844E+01)	1.029E+01 (8.866E-01)	6.230E-06 (3.100E-06)
F03	**3.854E-14 (2.063E-13)**	6.698E+03 (5.920E+01)	4.614E+03 (4.234E+02)	5.200E+03 (4.786E+02)	3.020E+04 (8.660E+02)	9.925E+02 (1.039E+02)	2.314E+03 (5.218E+01)	4.071E+04 (3.451E+03)	8.582E+03 (1.098E+03)	5.940E-02 (2.570E-02)
F04	**4.008E-13 (1.200E-12)**	9.540E+03 (1.450E+02)	6.501E+03 (4.488E+02)	9.530E+03 (4.141E+02)	8.900E+04 (1.140E+04)	4.163E+03 (7.877E+02)	5.894E+03 (3.232E+03)	2.997E+05 (4.337E+04)	1.844E+04 (3.928E+03)	1.570E+00 (8.780E-01)
F05	5.525E+03 (1.595E+02)	1.717E+03 (3.240E+01)	5.231E+02 (3.354E+01)	2.998E+03 (5.974E+02)	1.261E+03 (1.950E+02)	**3.672E+01 (8.211E+01)**	1.894E+02 (6.251E+01)	5.118E+03 (7.753E+02)	5.163E+02 (1.181E+02)	6.690E+03 (1.200E+02)
F06	**1.296E-14 (1.833E-14)**	8.065E+02 (9.232E+00)	1.479E+02 (1.880E+01)	3.105E+02 (3.235E+01)	4.510E+03 (1.070E+02)	3.577E+01 (5.306E+00)	9.184E+01 (6.196E+01)	8.800E+03 (2.698E+02)	1.955E+03 (2.053E+02)	9.880E-04 (1.880E-04)
F07	**6.387E-20 (6.822E-20)**	3.521E+03 (1.025E+02)	6.065E+02 (9.028E+01)	2.714E+03 (6.170E+02)	2.700E+04 (1.340E+03)	6.596E+02 (5.668E+01)	1.197E+03 (1.080E+02)	2.793E+04 (3.516E+03)	8.944E+03 (5.559E+02)	1.170E-12 (8.400E-13)
F08	**4.488E-13 (1.381E-12)**	8.049E+06 (3.006E+05)	1.618E+08 (1.730E+08)	1.162E+06 (3.312E+05)	2.800E+08 (1.560E+07)	5.345E+04 (1.039E+04)	1.105E+06 (1.497E+06)	2.220E+08 (6.795E+07)	6.887E+07 (1.728E+07)	4.550E-01 (1.720E-01)
F09	3.243E+03 (4.700E-02)	7.694E+06 (3.179E+05)	3.093E+05 (2.981E+04)	1.232E+06 (2.540E+05)	2.850E+08 (1.610E+07)	2.146E+04 (2.188E+03)	1.617E+05 (1.499E+05)	2.474E+08 (6.342E+07)	3.375E+07 (1.523E+07)	**3.240E+03 (2.110E-02)**
F10	**3.868E-11 (1.545E-10)**	2.348E+07 (1.098E+06)	1.783E+06 (3.731E+05)	2.525E+06 (3.979E+05)	2.360E+08 (3.430E+07)	2.979E+05 (3.386E+04)	1.856E+06 (2.150E+05)	4.912E+07 (1.306E+07)	4.512E+07 (4.660E+06)	8.250E+00 (2.830E+00)
F11	**1.723E-13 (2.761E-13)**	7.718E+02 (2.429E+01)	2.987E+03 (2.798E+02)	1.464E+03 (3.408E+02)	2.460E+04 (1.390E+03)	2.153E+02 (1.828E+01)	4.396E+03 (2.287E+02)	8.213E+03 (8.455E+02)	4.686E+03 (3.964E+02)	1.460E-01 (9.680E-02)
F12	**1.900E-08 (4.210E-08)**	5.240E+09 (1.249E+08)	8.774E+08 (1.589E+08)	1.831E+09 (3.251E+08)	2.280E+11 (4.240E+10)	9.847E+07 (1.554E+07)	1.441E+09 (2.555E+08)	6.290E+10 (1.579E+10)	1.493E+10 (3.024E+09)	1.040E+03 (5.480E+02)
F13	**2.301E-06 (1.568E-06)**	2.134E+03 (2.177E+01)	8.314E+02 (5.655E+01)	1.065E+03 (1.614E+02)	4.890E+03 (3.960E+01)	3.280E+02 (2.881E+01)	5.579E+02 (3.335E+02)	7.670E+03 (3.622E+02)	2.616E+03 (2.250E+02)	7.700E-01 (2.650E-01)
F14	**4.966E-09 (4.157E-09)**	3.904E+01 (1.084E+00)	9.796E+00 (1.219E+00)	1.959E+01 (1.081E+00)	6.180E+02 (1.890E+01)	1.625E+00 (3.524E-01)	1.461E+01 (4.553E+00)	4.814E+02 (8.168E+01)	1.411E+02 (1.137E+01)	1.880E-03 (3.870E-04)
F15	**1.643E-16 (2.176E-16)**	6.742E+03 (5.441E+01)	6.155E+03 (1.517E+02)	5.745E+03 (3.412E+02)	2.500E+04 (2.190E+03)	1.507E+03 (1.135E+02)	2.474E+03 (2.006E+02)	3.541E+04 (3.222E+03)	1.046E+04 (8.008E+02)	6.630E-02 (2.060E-02)
F16	5.777E+01 (1.919E+00)	6.194E+01 (1.262E+00)	6.294E+01 (5.167E-01)	3.822E+01 (2.995E+00)	7.600E+01 (2.290E+00)	5.498E+01 (7.596E+00)	**5.292E+00 (4.152E-01)**	5.106E+01 (1.001E+01)	6.135E+01 (8.279E-01)	6.210E+01 (1.030E+00)
F17	**9.866E-09 (7.606E-09)**	4.834E+00 (6.261E-02)	3.754E+00 (4.324E-01)	5.654E+00 (2.285E-01)	4.110E+02 (8.710E+01)	1.646E+00 (7.809E-02)	3.462E+00 (3.903E-01)	2.901E+01 (4.837E+00)	1.005E+01 (2.493E-01)	1.720E-03 (3.830E-04)
F18	**6.599E-08 (5.840E-08)**	1.884E+01 (3.008E-01)	1.362E+01 (3.275E-01)	2.357E+01 (1.766E+00)	3.770E+02 (4.290E+01)	6.241E+00 (4.583E-01)	1.104E+01 (1.035E+00)	9.710E+01 (6.117E+00)	4.182E+01 (1.232E+00)	9.270E-03 (1.510E-03)
F19	**2.504E-01 (2.661E-06)**	4.827E+01 (1.566E+00)	1.363E+01 (6.999E-01)	3.757E+01 (5.729E+00)	1.450E+03 (1.780E+02)	1.298E+01 (6.760E-01)	7.663E+00 (1.828E+00)	6.042E+03 (4.227E+02)	2.226E+02 (8.881E+01)	5.870E-01 (8.560E-02)
F20	**-1.554E+01 (7.974E+00)**	6.241E+04 (4.464E+03)	4.083E+03 (1.410E+03)	2.172E+05 (5.159E+04)	2.580E+06 (1.770E+05)	5.656E-01 (4.247E+00)	3.012E+04 (2.961E+04)	2.278E+05 (1.276E+05)	3.206E+05 (8.408E+04)	5.150E+00 (1.890E+00)
F21	8.145E+01 (2.937E-01)	2.071E+01 (1.521E+00)	1.035E+01 (2.383E+00)	1.539E+01 (1.940E+00)	1.110E+02 (8.810E+00)	7.721E+00 (7.265E-01)	**5.995E+00 (1.645E+00)**	8.648E+01 (4.499E-02)	3.202E+01 (3.005E+00)	8.280E+01 (4.420E-02)
F22	8.178E+01 (3.128E-01)	2.360E+01 (3.481E-01)	1.284E+01 (9.817E-01)	1.509E+01 (3.052E+00)	1.140E+02 (8.090E+00)	1.175E+01 (1.301E+00)	**7.622E+00 (1.248E+00)**	8.652E+01 (5.223E-02)	3.337E+01 (8.495E-01)	8.320E+01 (5.990E-02)
F23	1.598E+00 (2.807E-02)	1.591E+00 (3.908E-02)	1.606E+00 (4.583E-02)	1.581E+00 (3.355E-02)	3.820E+01 (1.420E+01)	1.629E+00 (4.858E-02)	**7.443E-02 (9.713E-03)**	1.371E+00 (2.827E-01)	1.595E+00 (2.838E-02)	1.160E+00 (1.980E-01)
F24	7.157E+03 (6.957E+01)	7.061E+03 (1.984E+02)	7.282E+03 (1.456E+02)	9.616E+03 (1.109E+03)	3.410E+06 (6.660E+05)	5.681E+03 (6.482E+01)	**3.131E+03 (2.774E+02)**	3.025E+04 (1.985E+03)	9.213E+03 (3.019E+02)	7.470E+03 (5.850E+01)
GIFT-Evo W/NS/L	–	19/2/3	19/2/3	19/1/4	N/A	18/2/4	18/0/6	21/3/0	20/1/3	23/0/1

Table 16: Terminal BBOB error under method-specific native budgets at $d = 1000$. Entries are mean (standard deviation); lower is better. GIFT-Evo uses 3,100 FEs, whereas ES, DE, PSO, L-SHADE, LES, LGA, and ABOM use 310,000 FEs. CMA-ES and B2Opt are explicitly N/A columns because their implementations could not be executed at this dimension; they are excluded from every comparison. The final row gives Holm-corrected $\mathrm{W/NS/L}$ counts under this asymmetric-budget diagnostic; it is not a common-budget comparison.

F	GIFT-Evo	ES	DE	PSO	CMA-ES	L-SHADE	LES	LGA	ABOM	B2Opt
F01	**5.469E-17 (9.664E-17)**	1.063E+03 (1.061E+01)	3.315E+03 (4.038E+03)	5.672E+02 (8.029E+01)	N/A	5.183E+01 (8.265E+00)	2.246E+02 (8.454E+01)	1.279E+04 (1.785E+02)	4.896E+03 (2.701E+02)	N/A
F02	**3.735E-18 (6.715E-18)**	4.834E+01 (2.370E+00)	1.004E+00 (2.568E-01)	7.252E+00 (4.597E+00)	N/A	6.107E-01 (1.527E-01)	4.647E+00 (1.993E+00)	7.047E+02 (7.831E+01)	1.156E+02 (1.827E+01)	N/A
F03	**1.356E-15 (3.976E-15)**	1.428E+04 (9.019E+01)	6.005E+03 (4.519E+02)	1.355E+04 (2.359E+02)	N/A	2.867E+03 (2.130E+02)	7.778E+03 (3.223E+02)	9.226E+04 (1.700E+03)	4.268E+04 (3.824E+03)	N/A
F04	**3.278E-12 (3.774E-12)**	2.177E+04 (2.021E+02)	1.260E+04 (8.580E+02)	2.739E+04 (2.805E+03)	N/A	1.413E+04 (1.366E+03)	1.826E+04 (5.338E+03)	9.784E+05 (5.312E+04)	9.102E+04 (7.459E+03)	N/A
F05	1.333E+04 (1.595E+02)	4.178E+03 (3.571E+01)	6.782E+02 (4.950E+01)	7.984E+03 (5.258E+02)	N/A	**2.085E+00 (4.662E+00)**	1.399E+03 (1.228E+02)	1.124E+04 (1.863E+03)	3.234E+03 (2.878E+02)	N/A
F06	**1.014E-14 (2.000E-14)**	1.774E+03 (5.705E+01)	3.930E+02 (4.735E+00)	9.323E+02 (3.611E+01)	N/A	1.073E+02 (7.851E+00)	7.496E+02 (4.006E+01)	1.660E+04 (1.264E+03)	9.125E+03 (2.240E+02)	N/A
F07	**8.114E-20 (9.977E-20)**	8.280E+03 (8.534E+01)	1.887E+03 (2.128E+02)	7.860E+03 (4.035E+02)	N/A	2.031E+03 (1.534E+02)	3.952E+03 (2.225E+02)	8.457E+04 (1.200E+04)	4.246E+04 (2.248E+03)	N/A
F08	**3.813E-12 (8.207E-12)**	8.331E+07 (5.223E+06)	2.748E+09 (3.748E+06)	4.837E+07 (9.457E+06)	N/A	8.563E+05 (1.542E+05)	3.591E+07 (2.745E+07)	2.414E+09 (3.623E+08)	2.431E+09 (3.394E+08)	N/A
F09	**6.493E+03 (3.832E-02)**	8.287E+07 (3.729E+06)	4.665E+06 (5.274E+05)	2.953E+07 (1.126E+06)	N/A	3.351E+05 (8.688E+04)	1.067E+07 (2.836E+06)	2.916E+09 (2.236E+08)	2.981E+09 (7.470E+08)	N/A
F10	**1.433E-10 (2.298E-10)**	5.770E+07 (1.593E+06)	3.305E+06 (4.054E+05)	7.685E+06 (7.889E+05)	N/A	7.969E+05 (8.446E+04)	6.066E+06 (5.889E+05)	2.732E+08 (7.812E+07)	2.497E+08 (1.306E+07)	N/A
F11	**3.183E-13 (6.994E-13)**	1.566E+03 (2.229E+01)	4.909E+03 (2.031E+02)	2.839E+03 (4.079E+02)	N/A	3.655E+02 (3.427E+01)	8.707E+03 (1.964E+02)	1.526E+04 (1.401E+03)	9.487E+03 (6.796E+02)	N/A
F12	**2.027E-08 (5.295E-08)**	1.235E+10 (2.133E+08)	2.505E+09 (5.292E+08)	7.692E+09 (1.319E+09)	N/A	5.411E+08 (7.643E+07)	7.368E+09 (8.870E+08)	1.947E+11 (4.227E+10)	1.322E+11 (5.229E+09)	N/A
F13	**4.744E-06 (3.477E-06)**	3.233E+03 (2.366E+01)	4.687E+03 (3.759E+03)	2.417E+03 (1.229E+02)	N/A	7.131E+02 (1.389E+01)	1.428E+03 (3.018E+02)	1.131E+04 (8.416E+01)	7.323E+03 (7.020E+02)	N/A
F14	**4.748E-09 (5.091E-09)**	6.803E+01 (5.982E-01)	1.707E+01 (9.502E-01)	4.816E+01 (5.886E+00)	N/A	3.835E+00 (3.979E-01)	5.199E+01 (5.472E+00)	9.484E+02 (9.562E+01)	5.331E+02 (2.983E+01)	N/A
F15	**8.984E-16 (1.866E-15)**	1.436E+04 (1.119E+02)	1.329E+04 (3.185E+02)	1.414E+04 (7.987E+02)	N/A	3.751E+03 (1.649E+02)	8.638E+03 (6.667E+02)	8.978E+04 (7.659E+03)	4.007E+04 (2.602E+03)	N/A
F16	6.362E+01 (1.346E+00)	6.622E+01 (9.506E-01)	6.598E+01 (9.442E-01)	4.580E+01 (3.248E+00)	N/A	5.593E+01 (6.939E-01)	**9.079E+00 (6.865E-01)**	5.658E+01 (7.397E+00)	6.663E+01 (5.532E-01)	N/A
F17	**1.410E-08 (1.190E-08)**	5.314E+00 (3.920E-02)	4.232E+00 (4.402E-01)	6.410E+00 (3.001E-01)	N/A	1.761E+00 (1.599E-02)	5.206E+00 (1.386E-01)	3.432E+01 (3.976E+00)	1.920E+01 (1.401E+00)	N/A
F18	**5.003E-08 (5.133E-08)**	2.089E+01 (1.311E-01)	1.685E+01 (1.002E+00)	2.531E+01 (1.035E+00)	N/A	7.264E+00 (3.678E-01)	1.952E+01 (1.517E+00)	1.268E+02 (1.147E+01)	6.911E+01 (1.464E+00)	N/A
F19	**2.504E-01 (6.791E-07)**	2.203E+02 (4.204E+00)	2.707E+01 (9.283E-01)	1.375E+02 (2.743E+01)	N/A	2.150E+01 (1.148E+00)	7.670E+01 (1.785E+01)	2.801E+04 (1.521E+03)	7.568E+03 (1.678E+03)	N/A
F20	**-8.123E+00 (6.435E+00)**	2.007E+05 (6.331E+03)	4.391E+04 (7.961E+03)	7.069E+05 (1.397E+05)	N/A	1.984E+03 (9.724E+02)	1.292E+05 (6.453E+04)	8.938E+05 (1.374E+05)	3.765E+06 (7.398E+05)	N/A
F21	8.289E+01 (1.129E-01)	2.336E+01 (1.811E-01)	2.090E+01 (2.307E+00)	2.786E+01 (4.720E+00)	N/A	1.620E+01 (2.379E+00)	**1.255E+01 (5.692E+00)**	8.655E+01 (7.115E-03)	7.590E+01 (1.498E+00)	N/A
F22	8.374E+01 (7.018E-02)	4.248E+01 (2.308E-01)	3.876E+01 (3.417E+00)	4.504E+01 (3.815E+00)	N/A	3.198E+01 (1.319E+00)	**3.174E+01 (8.073E+00)**	8.654E+01 (2.529E-02)	7.815E+01 (1.874E+00)	N/A
F23	5.829E-01 (7.747E-03)	5.802E-01 (5.678E-03)	5.783E-01 (6.001E-03)	5.659E-01 (6.729E-03)	N/A	5.506E-01 (2.096E-02)	**4.666E-02 (5.948E-03)**	5.314E-01 (8.753E-02)	5.790E-01 (6.669E-03)	N/A
F24	1.495E+04 (1.230E+02)	1.507E+04 (4.640E+01)	1.661E+04 (3.472E+02)	2.261E+04 (1.220E+03)	N/A	1.250E+04 (2.587E+02)	**9.687E+03 (7.510E+02)**	6.509E+04 (1.024E+03)	3.090E+04 (9.775E+02)	N/A
GIFT-Evo W/NS/L	–	19/2/3	20/1/3	19/0/5	N/A	18/0/6	18/0/6	21/3/0	20/1/3	N/A