Supplementary Material
From Importance Shifts to Landscape Topology: Characterizing and Exploiting Hyperparameter Spaces in Parallel Reinforcement Learning

Yingjie Zou and Zhong Fan*

University of Exeter, Exeter, UK
richiezou@foxmail.com, Z.Fan@exeter.ac.uk

This document collects extended definitions, full tables, robustness analyses, and a representation/plasticity study that support the main paper. All numbers are computed from the raw run corpus (37,200 PPO runs) with the regeneration scripts provided; fitness–distance correlation (FDC) values use the GraphFLA computation described in Section S4.

S1 GraphFLA Landscape-Feature Definitions

We compute 19 features with GraphFLA on the directed landscape graph (nodes = configurations; an edge points from a configuration to a fitter Hamming-1 neighbour). *Ruggedness*: local-optima (LO) ratio (fraction of nodes with no fitter neighbour); range-over-sill (R/S) ratio (fitness range over the variogram sill); autocorrelation (lag-1 along random walks); neighbour fitness correlation (Pearson correlation between a node's fitness and the mean fitness of its neighbours). *Epistasis* (proportions over HP pairs): magnitude, sign, reciprocal-sign, positive, and negative epistasis. *Navigability*: FDC (Spearman correlation between fitness and graph distance to the global optimum); global-optimum (GO) accessibility (fraction of nodes that can reach the global optimum via a monotonically improving path). *Neutrality*: fraction of neutral edges. *Fitness distribution*: skewness, kurtosis, coefficient of variation (CV), quartile coefficient, median-to-mean ratio, relative range, and Cauchy location.

S2 Experimental Details

Fixed HP defaults. Outside each chapter's swept HPs, all runs use the defaults in Table S1.

Continuous optimisers on a discrete grid. CMA-ES, DE, PSO, and (1+1)-ES propose continuous points; each proposal is snapped to the nearest grid point per ordinal dimension and looked up in the tabular surrogate. Repeated proposals re-use the cached value but still consume one evaluation, equalising budgets across methods.

* Corresponding author.

Table S1: Fixed HP defaults (used outside each chapter's swept HPs).

HP	Value	HP	Value
`discount`	0.99	`minibatch_size`	128
`gae_lambda`	0.95	`reuse_rollout_epochs`	4
`lr`	3×10^{-4}	actor/critic hidden	$[256, 256]$
`clip_epsilon`	0.2	`normalize_obs`	True
`grad_clip_norm`	1.0	`normalize_gae`	True
`actor_entropy`	-0.01	`critic_loss_weight`	0.5
`gae_horizon`	0		

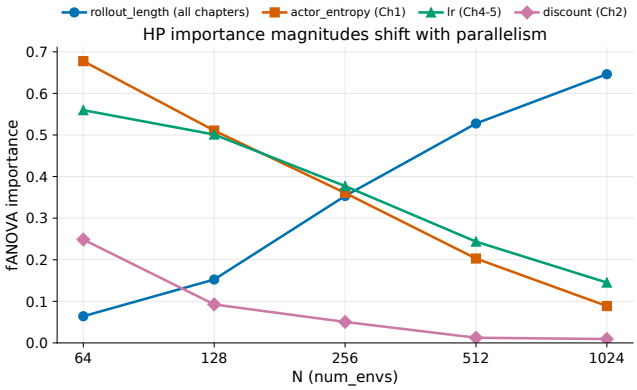

Fig. S1: Mean fANOVA importance vs. parallelism, pooled over the chapters and environments in which each HP is swept. `rollout_length` rises **10.1**× (**0.064** → **0.646**); the discount factor falls **26.6**× in Chapter 2.

S3 HP Importance: Full Results

Figure S1 shows the magnitude shifts behind the headline 10.1×. Figure S2 gives the per-HP max/min ratio across all chapters, Fig. S3 the full 5×5 chapter×regime fANOVA heatmap, Fig. S4 PED-ANOVA gating, and Fig. S5 the stage-wise evolution.

S4 Landscape Topology: Full Table and Block-Robust Statistics

Figure S6 shows all 19 features vs. parallelism; Fig. S7 the per-chapter profiles. Because the 100 per-regime slices form 20 chapter–environment blocks × 5 levels, Table S2 reports a block-respecting trend test: the Spearman ρ computed *within* each block and aggregated across blocks with a Wilcoxon signed-rank test. Every headline trend retains its direction and significance.

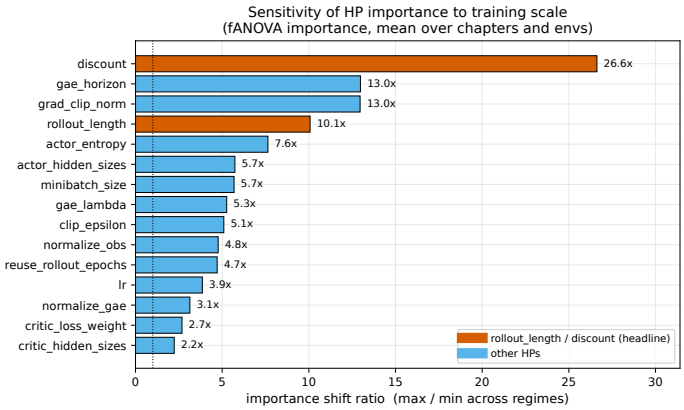

Fig. S2: Per-HP importance shift ratio (max/min across parallelism levels).

Table S2: Block-structure-robust trend test (R1-8). "within-block ρ" is the mean per-condition Spearman correlation vs. `num_envs` over the 20 chapter–environment blocks; p is a Wilcoxon signed-rank test on the 20 per-block correlations (independence-respecting, unlike the pooled p).

Feature	pooled ρ	within-block ρ	robust p
LO ratio	-0.564	-0.714	< 0.001
R/S ratio	-0.316	-0.535	0.005
Neighbour fit. corr.	$+0.362$	$+0.445$	0.022
Reciprocal sign epistasis	-0.681	-0.885	< 0.001
FDC	-0.265	-0.430	< 0.001
GO accessibility	$+0.563$	$+0.735$	< 0.001
Skewness	$+0.611$	$+0.875$	< 0.001
CV	$+0.685$	$+0.985$	< 0.001

S5 Algorithm Selection: Full Results

Figure S8 reports all eight methods across four budgets; Table S3 the feature–method correlation matrix (Fig. S9 as a heatmap); Table S4 the per-landscape rankings; Fig. S10 the selector baselines at n=20. Figure S11 shows the n=100 per-regime selector. Because the eight optimisers are stochastic and the leading methods sit within ≈0.003 regret, the per-slice oracle-best label is seed-dependent; over 10 seeds the feature-based LOO accuracy is 23±8% (range 10–36%; vs. 5% at n=20 and 12.5% uniform) but does not reliably exceed the majority-class baseline (26±3%, winning in only 4/10 seeds). Feature-based meta-modelling thus remains a negative result even at 5× the data.

Table S3: Spearman correlations between landscape features and method regret at B=40 (n=20 conditions). $^{*}/^{**}/^{***}$: $p < 0.05/0.01/0.001$ uncorrected; †: significant after Holm–Bonferroni correction (152 tests). Only features with ≥1 corrected-significant correlation shown.

Feature	RS	DE	CMA-ES	GA	PSO	RMHC	SA	OnePlusOneES
rel_range	**+0.70*****	**+0.73**†	**+0.82**†	**+0.61****	+0.47*	**+0.74**†	**+0.73**†	**+0.55***
skewness	**+0.65****	**+0.69*****	**+0.80**†	**+0.61****	+0.42	**+0.73**†	**+0.76**†	**+0.55***
med_mean	**-0.67****	**-0.67****	**-0.68*****	**-0.62****	-0.38	**-0.77**†	**-0.75**†	**-0.62****
ep_neg	-0.42	**-0.53***	**-0.56***	**-0.51***	-0.28	**-0.63****	**-0.72**†	-0.49*
ep_pos	+0.42	**+0.53***	**+0.56***	**+0.51***	+0.28	**+0.63****	**+0.72**†	+0.49*
lo_ratio	-0.50*	-0.48*	**-0.71*****	-0.31	-0.49*	-0.26	-0.39	-0.26
$C\bar{V}$	+0.45*	**+0.52***	**+0.66****	+0.44	+0.32	**+0.51***	**+0.57****	+0.31

Table S4: Mean normalized regret per landscape type (chapter) and method at B=40. **Bold**: best per row. Friedman test across 20 conditions: χ^2=30.0, $p < 0.001$.

Landscape	RS	DE	CMA-ES	GA	PSO	RMHC	SA	OnePlusOneES
Chap1	0.114	0.109	0.097	0.089	**0.068**	0.128	0.111	0.109
Chap2	0.073	0.065	**0.054**	0.065	0.067	0.073	0.062	0.074
Chap3	0.075	0.083	**0.063**	0.068	0.093	0.111	0.078	0.071
Chap4	0.105	0.088	0.086	0.082	**0.077**	0.111	0.096	0.090
Chap5	0.083	0.084	0.082	0.081	**0.070**	0.092	0.087	0.091
Overall	0.090	0.086	0.076	0.077	**0.075**	0.103	0.087	0.087
Mean Rank	5.6	4.4	3.4	3.5	**3.1**	6.3	4.8	4.8

S6 Representation and Optimisation Diagnostics

Following Mayor et al., we probe network-level effects of parallelism on the trained PPO policies (Chapter 1; 4 environments × 5 parallelism levels × 3 seeds, default HP configuration, final checkpoint), computing four metrics directly from the policies via the evorl rollout pipeline (Fig. S12).

- **Dormant neurons**: scoring each hidden unit by its mean post-ReLU activation on real (preprocessed) observations, normalised by the layer mean, a mean of 18% of units are dormant at the Sokar threshold τ=0.025 (12% strictly dead), strongly environment-dependent (up to 51%), with *no* significant parallelism trend (ρ=+0.19, p=0.16). Moderate dormancy is therefore present—as on Atari—but is not driven by parallelism in this continuous-control setting.
- **Gradient heavy-tailedness**: the per-component excess kurtosis of the PPO actor-loss gradients *decreases* with parallelism (ρ=−0.46, p<0.001)—more parallel data yields more stable, less heavy-tailed gradients.
- **Policy / coverage**: the policy action standard deviation rises mildly (ρ=+0.36, p=0.005) and the participation ratio of visited observations rises weakly (ρ=+0.22, p=0.09).

The clearest network-level imprint of parallelism is thus *reduced gradient heavy-tailedness*; dormancy and coverage shift little. These network-level effects are modest relative to the structural shifts in the HP optimisation problem documented in the main paper.

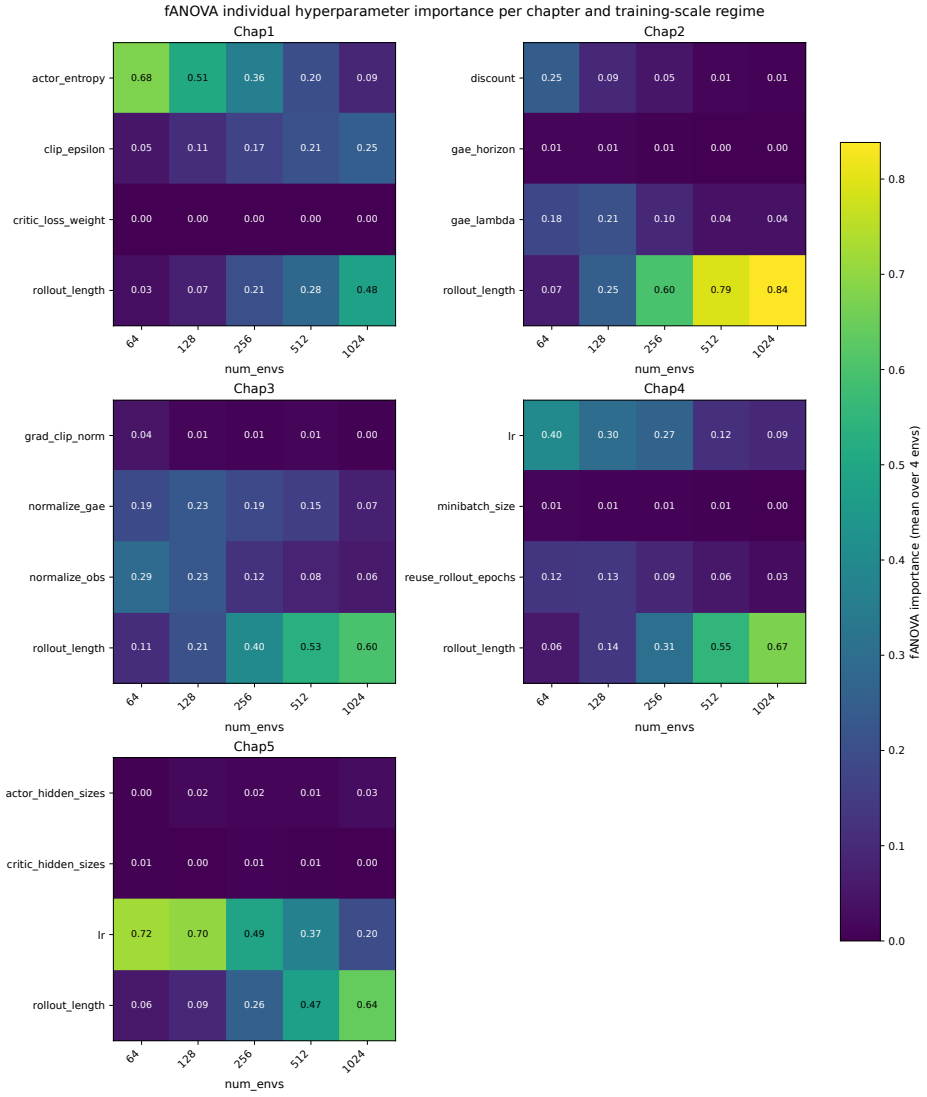

Fig. S3: fANOVA individual-importance heatmaps (5 chapters × 5 parallelism levels, mean over environments).

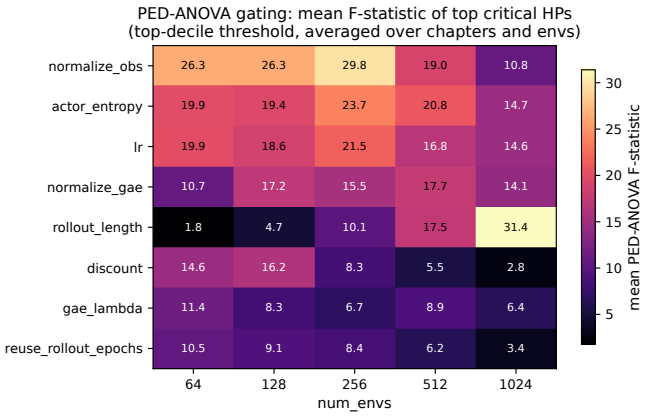

Fig. S4: PED-ANOVA gating: top critical-HP F-statistics.

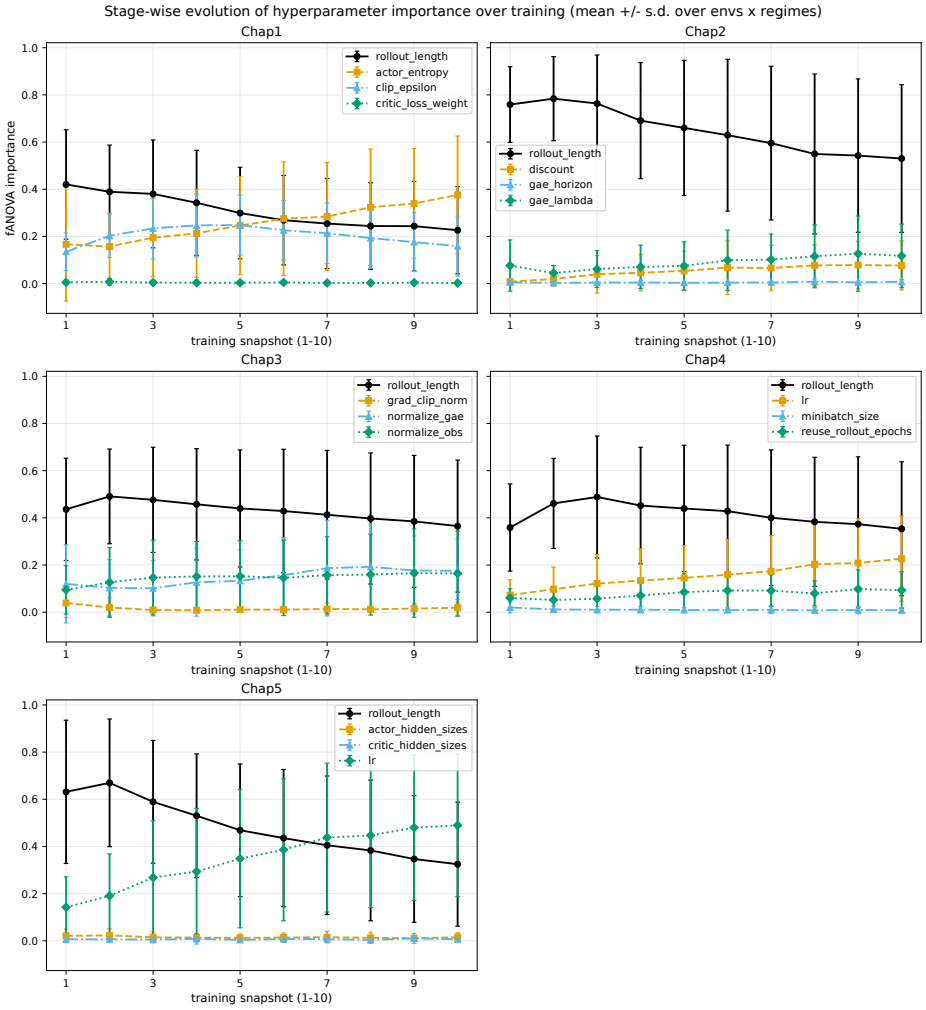

Fig. S5: Stage-wise importance over 10 training snapshots, per chapter.

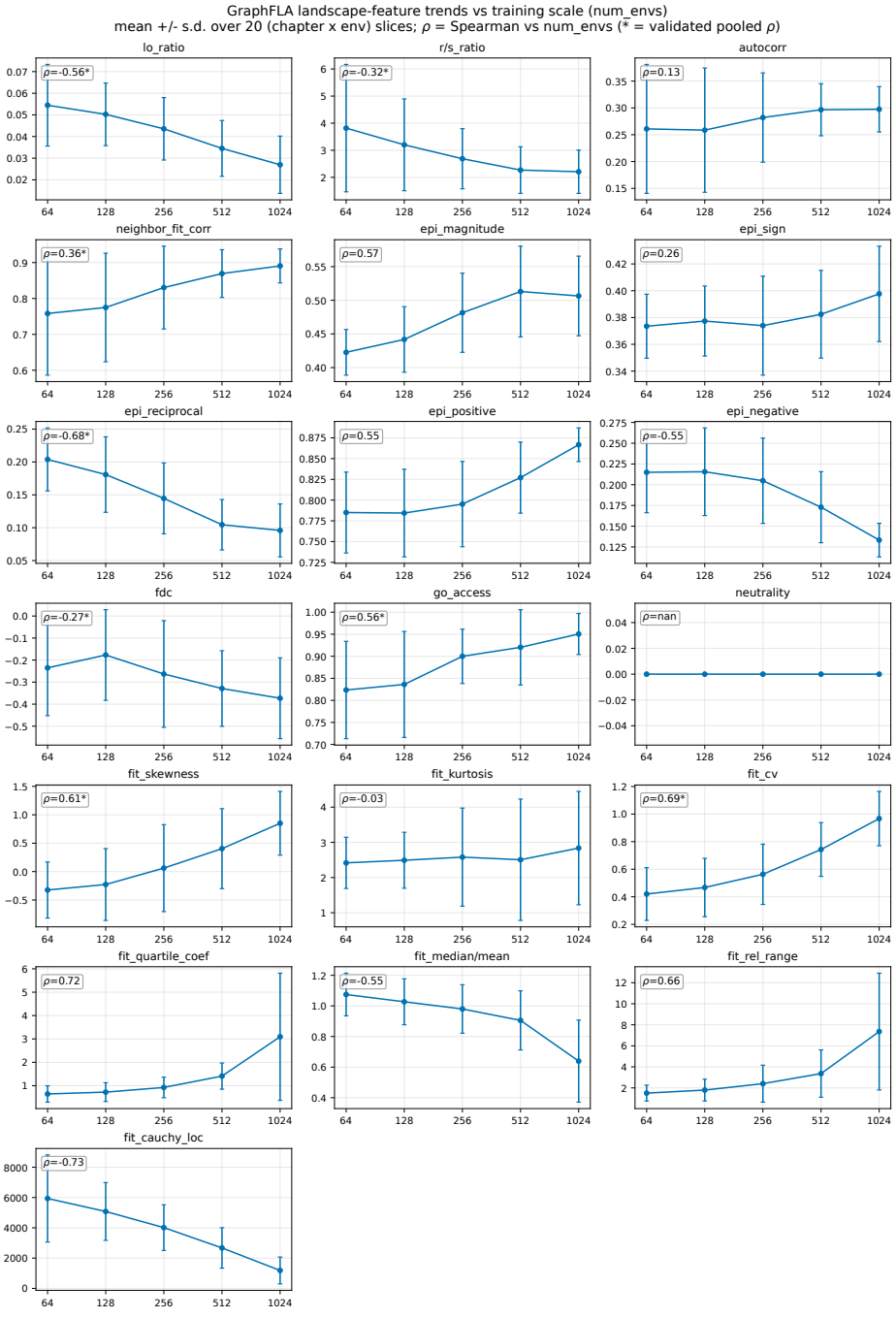

Fig. S6: All 19 GraphFLA features vs. parallelism (per-level means with Spearman ρ).

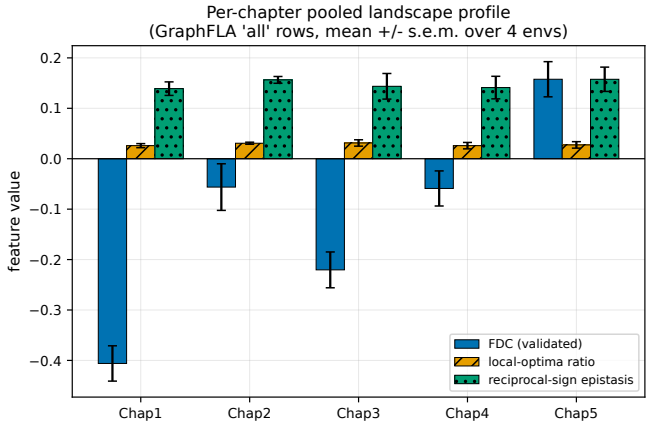

Fig. S7: Per-chapter landscape profiles (FDC, LO ratio, reciprocal-sign epistasis) on the pooled slices.

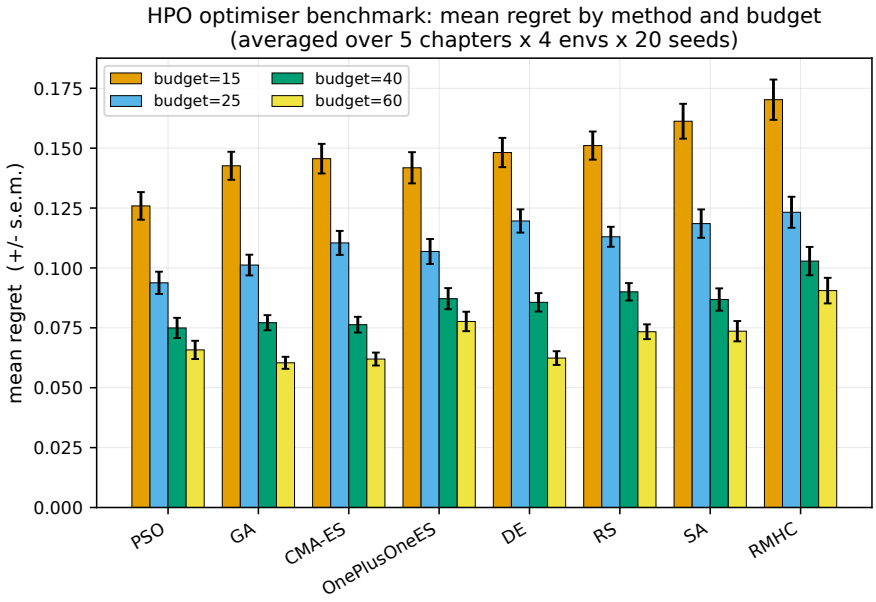

Fig. S8: Mean normalized regret, 8 methods × 4 budgets (±s.e.m.).

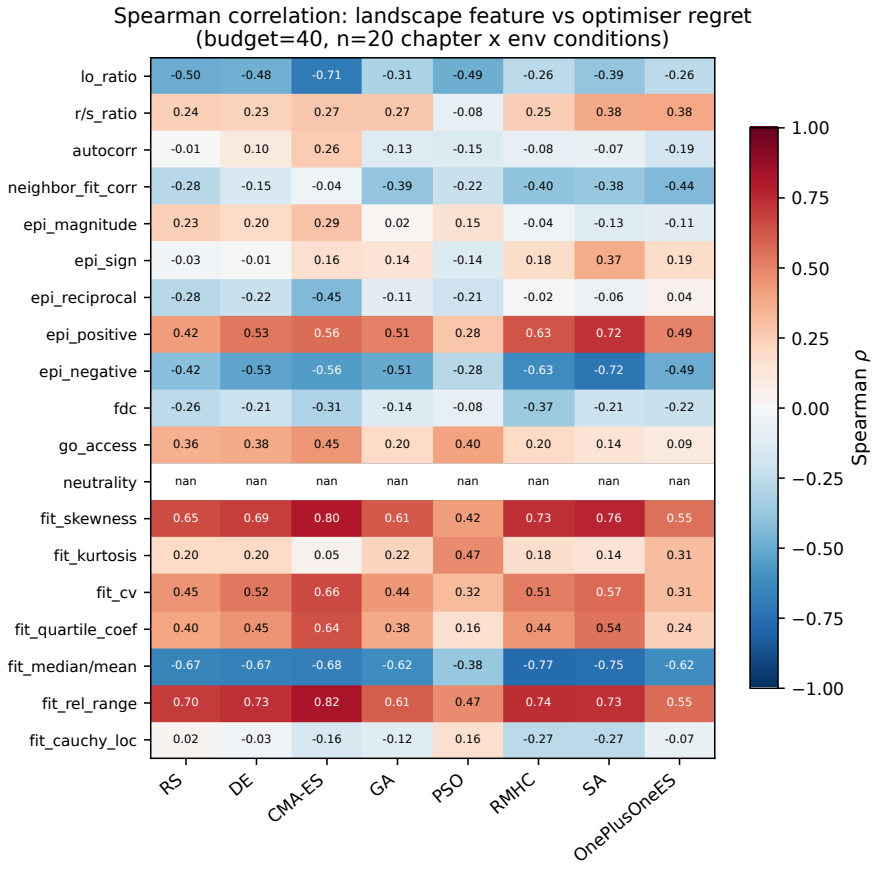

Fig. S9: Spearman correlations, 19 features × 8 methods, at B=40.

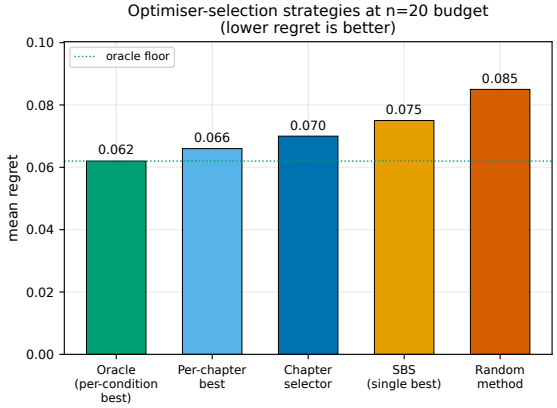

Fig. S10: Selection strategies at n=20 (oracle → random).

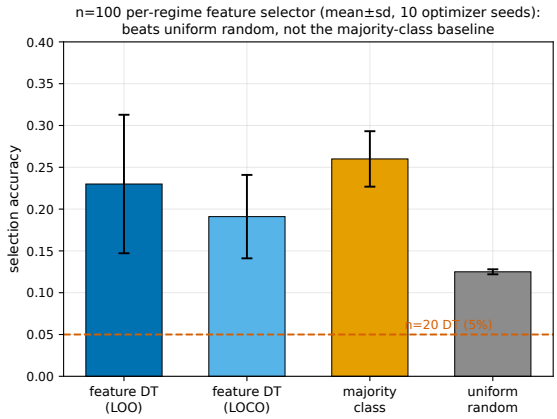

Fig. S11: Feature-based selector accuracy at n=100 vs. baselines.

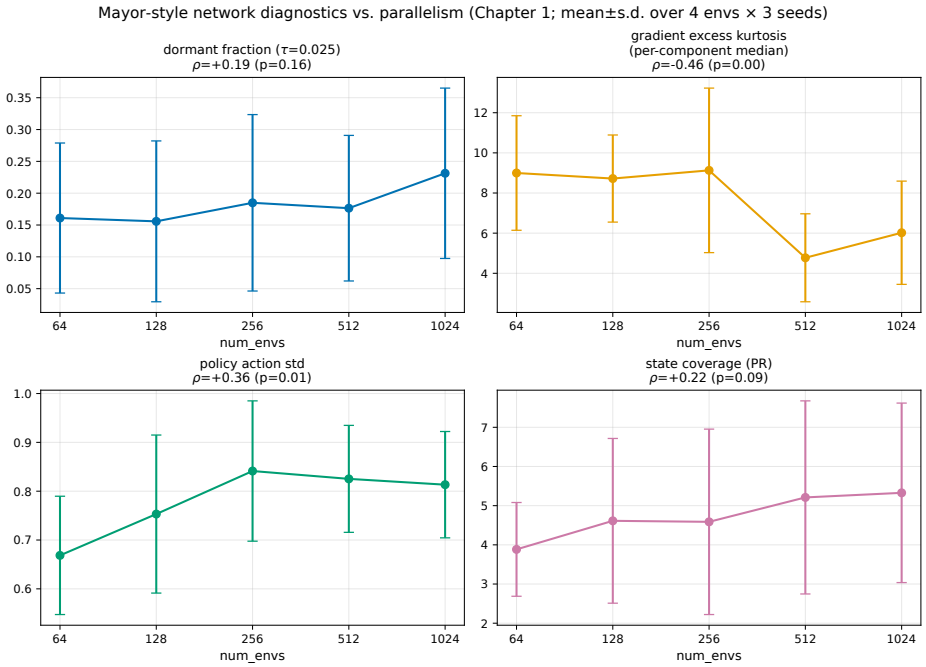

Fig. S12: Mayor-style network diagnostics vs. parallelism (Chapter 1; mean±s.d. over 4 environments × 3 seeds): dormant-neuron fraction (τ=0.025), per-component gradient excess kurtosis, policy action std, and state-coverage participation ratio. Gradient heavy-tailedness falls significantly with parallelism; dormancy and coverage shift little.