SCOPE: EXPANDING GENERATIVE COVERAGE FOR STELLARATOR DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Stellarator design for fusion requires multiple plasma configurations that satisfy requested properties and provide alternatives for subsequent physics and engineering development. Finding these configurations is expensive, and high-quality training designs are scarce. Generative models can accelerate candidate production, creating an opportunity to expand their repertoire by learning from new physical evaluations. This introduces a mismatch between discovery and learning: once a design enters the search archive, it ceases to be novel, while the generator may still need further training to learn its region. We introduce SCOPE, a framework that separates novelty for current exploration from the continued learning value of discoveries. SCOPE starts from a conditional flow model of full plasma boundaries trained on physically screened designs. In each round, generated candidates are checked for equilibrium, request satisfaction, and physical quality, and compared against a growing archive to assess novelty. This feedback guides policy updates toward novel, useful designs. We retain valid designs for further training and prioritize discoveries during replay. Experiments show that SCOPE generates higher-quality candidates and broadens exploration within the target design domain. It produces $2.4\times$ more distinct qualified designs beyond current baseline by direct sampling from the trained model. In online search, SCOPE is able to discover more valid and qualified designs under controlled budget.

1 INTRODUCTION

Stellarators confine plasma using shaped, three-dimensional magnetic fields, offering a path toward steady-state fusion energy. Their design combines searches over plasma configurations with subsequent physical refinement, coil design, and engineering selection. Design sweeps seek multiple configurations at requested values of compactness and magnetic-field twist, providing alternatives for these coupled stages (Giuliani et al., 2022; Bindel et al., 2023). For a conditional generator, a valid response must both fulfill the request and admit a magnetohydrodynamic (MHD) equilibrium. Among valid responses, quasisymmetry (QS), which supports particle confinement (Landreman & Paul, 2022), and magnetic-well quality, a complementary stability indicator, identify promising candidates for further development. Building this *repertoire of distinct, request-valid designs* is expensive: evaluating each candidate requires a numerical equilibrium solve, and boundary optimization repeats this calculation to improve quality. We study how a generator can expand this repertoire through successive rounds of physical feedback, with a fixed evaluation budget per round.

Conditional generative models can rapidly propose stellarator candidates (Padidar et al., 2025). Their sampling priorities are learned from finite repositories, and feedback concentrated on familiar, high-scoring configurations can reinforce the same regions. Continued exploration calls for learning from new physical evidence to search beyond an expanding archive. A central distinction is that *discovering a design and learning to generate its region are separate achievements* (Figure 1). Recording a new configuration immediately makes it familiar to the archive; generating its neighborhood under the same request requires the policy to absorb that discovery. Novelty for search and value for learning therefore evolve on different timescales. Search should reward new regions, while learning should continue to reinforce valid discoveries after archival. The challenge is to turn this physical feedback into reusable capability, enabling the model to *explore, learn, and explore again*.

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.

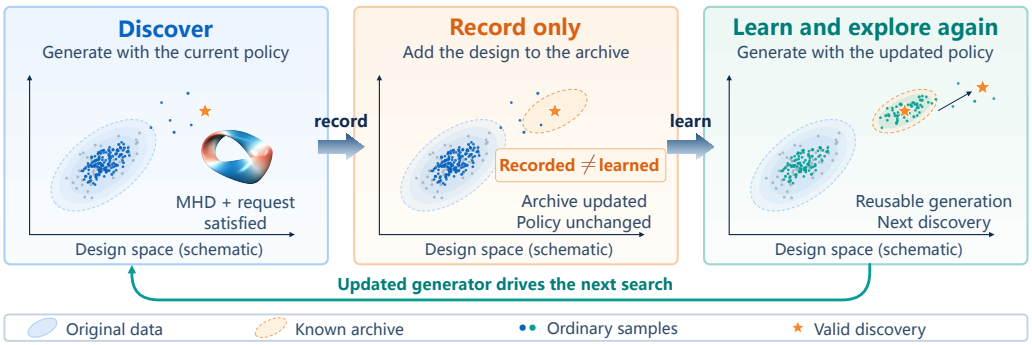


Figure 1: **Discovering a design and learning to generate its region are separate steps.** A request-valid equilibrium expands the known archive. Learning makes its region accessible to ordinary generation and supplies proposals for further exploration. Ordinary generation uses direct model samples, which undergo physical evaluation without preselection by novelty or quality. The middle panel illustrates an archive update with the policy held fixed. Design-space distributions are schematic.

We introduce SCOPE, a framework that turns physical discoveries into a generative policy for continued exploration. We first pretrain a conditional flow model on the full dataset and fine-tune it on designs selected for QS and magnetic-well quality. Starting from this model, each online round generates candidates and updates the policy using rewards based on MHD recovery, request fulfillment, physical quality, and novelty relative to a growing archive. To turn discoveries into reusable generation capability, we retain valid designs for replay and continue prioritizing discoveries after they enter the archive. Learning thus continues after novelty disappears, allowing each round's discoveries to support subsequent search and generative design sweeps.

The trained SCOPE generator meets the requested conditions and physical criteria in 65–71% of outputs across the evaluated request sets, compared with 9–10% for Diffusion for Fusion (DFF). In the target domain, it generates 2.4 times as many distinct qualified designs beyond the original corpus as DFF for the same number of evaluated samples. In a separate eight-round comparison with the same starting model and search budget, SCOPE finds 17% more qualified designs than retraining on valid candidates. Within SCOPE, giving discoveries replay priority for one round yields more discoveries during search, while maintaining that priority across rounds helps the final model generate more distinct designs.

Our contributions are:

- We prepare physically annotated stellarator data and train a conditional flow model directly in the full boundary coefficient space.
- We design online policy updates that reward new valid designs while continuing to learn from discoveries already recorded in the archive.
- We demonstrate higher design quality and continued discovery, and show how replay priority changes the balance between finding new designs and generating distinct designs afterward.

2 PRELIMINARIES AND PROBLEM FORMULATION

Plasma boundaries and equilibrium. A stellarator plasma has a shaped, toroidal boundary, periodic around the tube (poloidal angle ϑ) and the torus (toroidal angle φ). Truncated Fourier series represent its cylindrical radius and height, $R(\vartheta, \varphi)$ and $Z(\vartheta, \varphi)$. Their coefficients form the generated vector $x \in \mathbb{R}^d$: low-order modes describe broad shape and higher-order modes add fine structure. Given x and prescribed pressure and current profiles, magnetic flux, and dimensional scale, collectively η, an equilibrium solver balances magnetic forces against pressure gradients (Panici et al., 2023). We denote geometric admissibility and successful numerical MHD recovery by $R(x; \eta) = 1$. This establishes the physical basis for evaluating the design.

Design requests and physical quality. Requests specify aspect ratio A, the torus's major radius relative to its cross-sectional radius, and rotational transform ι, the number of poloidal field-line turns per toroidal circuit. These describe compactness and magnetic-field twist; we use the signed vacuum component of ι (Appendix A). Among equilibria meeting these targets, physical quality guides further selection. *Quasisymmetry* is a symmetry of magnetic-field strength that supports particle confinement in a three-dimensional configuration (Landreman & Paul, 2022); its normalized violation Q should be small. The *magnetic well* provides a complementary stability indicator (Cadena et al., 2025), with profile $W(\rho)$ on nested surfaces indexed by ρ. Low QS violation and nonnegative well throughout the evaluated interior identify promising candidates for subsequent design stages.

Conditional generation and task validity. From a finite repository $\mathcal{D}_0$, we learn to sample boundaries $x \sim p_\theta(\cdot \mid c)$ for targets $c = (A^\star, \iota^\star)$. The full request is $u = (c, \eta)$: the generator receives c, while η fixes the equilibrium problem throughout that request's search. A valid response, $V(x; u) = 1$, passes the physical checks $R(x; \eta) = 1$ and meets the requested condition tolerances. Joint qualification, $G_\tau(x; u) = 1$, further requires $Q < \tau$ and nonnegative $W(\rho)$ throughout the interior grid $\mathcal{P}$.

Exploration and reusable coverage. Under a physical-evaluation budget, we seek distinct request-valid designs and a policy that can sample their neighborhoods. QS and well guide exploration toward jointly qualified candidates. Diversity describes output variation; exploration measures novelty relative to accumulated knowledge. We evaluate ordinary coverage beyond $\mathcal{D}_0$ and discovery against an expanding archive, with V and G_τ specifying each outcome's physical value. Transfer to additional requests measures reuse across conditions. Section 3.3 defines geometric novelty.

Related work. QUASR provides optimized stellarator configurations (Giuliani et al., 2025), and Diffusion for Fusion learns conditional generation from this data (Padidar et al., 2025). Feedback-based design includes adaptive sampling in CbAS (Brookes et al., 2019) and advantage-weighted flow learning in Advantage Weighted Matching (AWM) (Xue et al., 2026). Quality-diversity methods build repertoires of high-performing solutions: learned representations can guide subsequent search (Gaier et al., 2020), and archive distillation makes repertoires accessible through a single policy (Faldor et al., 2023) or generative model (Hegde et al., 2023). For stellarator exploration, SCOPE couples equilibrium certification with request-conditioned generation. Current archive novelty directs search, while accumulated valid designs remain training targets. We evaluate both new discoveries and the repertoire available through ordinary generation.

3 SCOPE: LEARNING TO SEARCH BEYOND AN EXPANDING ARCHIVE

SCOPE combines full-boundary generation and joint-quality initialization to establish the proposal policy (Figure 2). *Archive-guided policy refinement* then closes the loop: each round proposes designs, evaluates request validity, quality, and novelty, and updates the policy that drives the next search.

3.1 A PHYSICALLY FAITHFUL BOUNDARY POLICY

Direct generation in boundary coefficient space. We generate plasma boundaries by modeling their Fourier coefficients directly. With stellarator symmetry and Fourier truncation $M = N = 6$, the representation contains 169 coefficients; fixing the major-radius scale leaves $d = 168$ free coordinates. This modest dimension makes it practical to retain every free coefficient, including small-amplitude harmonics, and train the flow model using fixed, invertible coordinate standardization. Table 1 supports this choice: direct-boundary FM yields higher joint physical qualification than the 24-dimensional Latent-FM baseline on the same offline data.

A conditional residual network represents $v_\theta(x_t, t, c)$, with time embeddings and request modulation at each block. Given a standardized boundary x and noise $\epsilon \sim \mathcal{N}(0, I)$, we interpolate as $x_t = (1 - t)\epsilon + tx$. Flow matching (Lipman et al., 2023) minimizes

$$e_\theta(x, c; t, \epsilon) = \frac{1}{d} \left\| v_\theta(x_t, t, c) - (x - \epsilon) \right\|_2^2, \qquad \ell_\theta(x, c) = \mathbb{E}_{t,\epsilon} e_\theta(x, c; t, \epsilon). \tag{1}$$

Integrating the velocity from noise defines $p_\theta(x \mid c)$. The matching objective supports both pretraining and online refinement, using the same full-boundary sampler throughout.

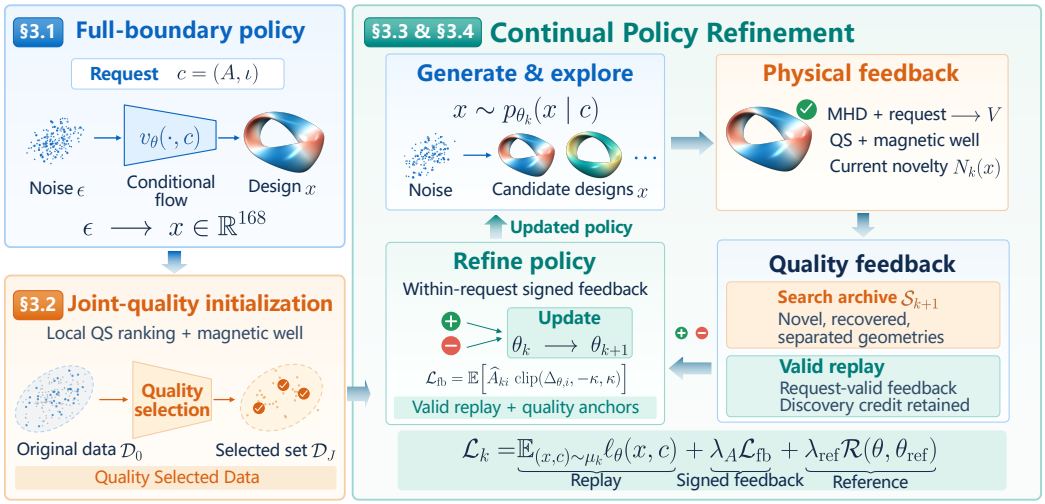


Figure 2: **A physical prior supports repeated exploration and policy learning.** Full-boundary flow and joint-quality initialization establish the proposal policy. The online loop evaluates generated designs and updates the policy with signed feedback and replay. Current novelty guides search, while request-valid replay and quality anchors carry useful designs into subsequent generation.

3.2 Joint-quality initialization and physical feedback

Establish a useful starting distribution. We pretrain on the broad corpus and refine on $\mathcal{D}_J$, which retains examples with nonnegative magnetic well and QS violation ranked within the best fraction q of their local request neighborhood. Local ranking adapts the QS cutoff to the quality distribution near each request (Appendix A.2). This *J-direct* stage supplies the quality prior and anchors for policy refinement (Figure 2).

Physical evaluation of generated designs. Each generated boundary is evaluated under its prescribed settings η for physical recovery, request fulfillment, QS, and magnetic well. Designs satisfying $V = 1$ supply positive replay examples; their continuous QS and well measurements guide the quality preference. We pair each accepted boundary with its generating request, so all responses within that request's tolerance region reinforce the condition they fulfill.

3.3 Reward discovery relative to accumulated knowledge

Each round generates more candidates than the physical budget permits evaluating. We form an evaluation batch $\mathcal{B}_k$ by reserving ordinary samples before geometric selection and prioritizing separated novel candidates for the remaining evaluations (Figure 2). Evaluation holds generated boundaries fixed and determines recovery R, request validity V, and joint qualification G_τ. In the controlled study, separated novel geometries with $R = 1$ enter the archive for the next round; their V/G_τ subsets measure discoveries meeting the request and quality criteria (Appendix A.5). All evaluated designs with $V = 1$ enter positive replay (Section 3.4).

Measure novelty against known designs. The archive $\mathcal{S}_k$ contains the initial reference and geometries accumulated across requests. We compare boundaries using full normal displacement, its low-frequency component, and dense surface-coordinate displacement. A candidate is novel, $N_k(x) = 1$, when its nearest distance under each metric d_j exceeds the fixed calibrated scale δ_j. Each metric independently searches the entire archive. Selection and scoring share one archive per round; the archive expands between rounds while the scales stay fixed.

Reward distinct, valid discoveries. For discovery rewards, we select representatives separately among candidates in $\mathcal{B}_k$ satisfying $V = 1$ and $N_k = 1$, using the same geometric separation rule as archive admission. We mark these representatives by $m_{ki} = 1$ and give them the discovery bonus,

so nearly identical proposals do not repeatedly earn credit for the same direction. To also reward progress toward the novelty thresholds, we use $s_k(x) = \min\{1, \min_j d_j(x, \mathcal{S}_k)/\delta_j\}$, which saturates when all three thresholds are reached. A utility $U(x)$ favors low QS violation and a larger fraction of positive magnetic-well nodes. Candidates failing the physical or request checks receive a penalty:

$$r_{ki} = \begin{cases} \lambda_N m_{ki} + \lambda_d s_k(x_i) + \lambda_Q U(x_i), & V(x_i; u_i) = 1, \\ -\lambda_{\text{fail}}, & V(x_i; u_i) = 0. \end{cases} \tag{2}$$

Utility definitions and weights appear in Appendix A.4.

For learning, we compare each candidate with other responses to the same request. Centering and scaling rewards by their current mean $\bar{r}_{k,c}$ and standard deviation $\sigma_{k,c}$ gives

$$\widehat{A}_{ki} = \text{clip}\left(\frac{r_{ki} - \bar{r}_{k,c_i}}{\max(\sigma_{k,c_i}, \sigma_{\min})}, -A_{\max}, A_{\max} \right). \tag{3}$$

Positive and negative advantages identify responses above and below the request's average reward.

3.4 Archive-guided policy refinement

Archive novelty changes as soon as a design is recorded; the policy learns its region through subsequent updates. We use current archive-relative rewards to direct search and fit accumulated valid designs to build reusable generation.

Learn from accumulated valid designs. We train on the valid designs collected during search, together with previously available high-quality examples that serve as quality anchors. Denoting their sampling distributions by $\mu_{\text{online},k}$ and μ_{quality}, respectively, we use

$$\mu_k = (1 - \gamma)\mu_{\text{quality}} + \gamma \mu_{\text{online},k}, \qquad 0 \le \gamma \le 1. \tag{4}$$

Replay balances requests. In the full configuration, stored acquisition markers m_{ki} give discovery representatives extra weight as the archive expands (Figure 2). Section 4.5 ablates replay allocation and priority duration; Appendix A.4 specifies the pools and weights.

Favor responses with higher search value. Current rewards guide fitting as follows: positive advantages encourage lower matching error, while negative advantages reverse that direction. This correction uses all evaluated responses, including those failing validity checks. For small $\xi > 0$, reweighting responses by $1 + \xi \widehat{A}_{ki}$ adds $\xi \mathbb{E}_i[\widehat{A}_{ki} \ell_\theta(x_i, c_i)]$ to the matching objective.

We implement this signed correction through clipped changes in error from the frozen policy $\bar{\theta}_k$ that collected the batch. Using the same interpolation state for both policies gives

$$\Delta_{\theta,i} = \frac{e_\theta(x_i, c_i; t, \epsilon) - e_{\bar{\theta}_k}(x_i, c_i; t, \epsilon)}{\text{sg}[\max(\bar{e}_{\bar{\theta}_k}, s_{\min})]}, \qquad \mathcal{L}_{\text{fb}}(\theta) = \mathbb{E}_{i,t,\epsilon}\left[\widehat{A}_{ki} \ \text{clip}(\Delta_{\theta,i}, -\kappa, \kappa) \right]. \tag{5}$$

Here $\bar{e}_{\bar{\theta}_k}$ is the old policy's mean minibatch error and sg stops gradients. Within the clipping range, the advantage sets the correction's direction; clipping bounds each sample's contribution.

Regularize changes to the initial policy. To keep refinement close to the initial generative behavior, we use the initial policy θ_{ref} as a fixed reference and penalize normalized velocity differences through $\mathcal{R}$. The resulting update fits accumulated valid designs, uses current rewards to favor more useful responses, and regularizes policy changes:

$$\mathcal{L}_k(\theta) = \underbrace{\mathbb{E}_{(x,c)\sim\mu_k} \ell_\theta(x, c)}_{\text{fit valid designs}} + \underbrace{\lambda_A \mathcal{L}_{\text{fb}}(\theta)}_{\text{use current feedback}} + \underbrace{\lambda_{\text{ref}} \mathcal{R}(\theta, \theta_{\text{ref}})}_{\text{regularize policy change}}. \tag{6}$$

During each update window, we also monitor drift from the rollout policy $\bar{\theta}_k$ and retain the last accepted iterate. Appendix A.4 specifies the scales and drift rule. Setting $\lambda_A = 0$ gives SCOPE's positive replay ablation (ELITE). Algorithm 1 combines these learning steps with updates to the search archive.

Algorithm 1 SCOPE: archive-guided policy refinement

Input: initial policy θ_0, archive $\mathcal{S}_0$, quality anchors, requests, K rounds, and physical budget per round.

$\theta_{\text{ref}} \leftarrow \theta_0$; $\mathcal{D}_{\text{replay}} \leftarrow \varnothing$.
for $k = 0, \ldots, K-1$ **do**
  $\bar{\theta}_k \leftarrow \theta_k$ (freeze the rollout policy).
  $\mathcal{C}_k \leftarrow$ generate candidates from $p_{\bar{\theta}_k}(x \mid c)$ for each request c.
  $\mathcal{B}_k \leftarrow$ select ordinary and frontier candidates from $\mathcal{C}_k$ using $\mathcal{S}_k$.
  Evaluate $\mathcal{B}_k$ for physical recovery, request validity, QS, and magnetic well.
  Compute N_k, representative markers m_{ki}, and rewards r_{ki} against $\mathcal{S}_k$.
  Normalize rewards within requests to obtain $\widehat{A}_{ki}$ (Equation 3).
  $\mathcal{D}_{\text{replay}} \leftarrow \mathcal{D}_{\text{replay}} \cup \{(x_i, c_i, m_{ki}) : i \in \mathcal{B}_k,\ V(x_i; u_i) = 1\}$.
  Form μ_k from anchors and replay, using stored markers (Equation 4).
  $\theta_{k+1} \leftarrow \text{Update}(\theta_k; \mu_k, \mathcal{B}_k, \{\widehat{A}_{ki}\}, \bar{\theta}_k, \theta_{\text{ref}})$ (Equation 6).
  $\mathcal{S}_{k+1} \leftarrow$ expand $\mathcal{S}_k$ under the archive-admission rule (Section 3.3).
end for
return policy θ_K and archive $\mathcal{S}_K$.

4 EXPERIMENTS

We first evaluate the quality and design coverage attained by the complete SCOPE system, then test its online learning design under controlled feedback and update budgets.

4.1 EXPERIMENTAL SETUP

Evaluation protocols. Tables 1 and 2 compare *end-to-end generation quality and coverage* under a common final evaluation budget. The complete SCOPE pipeline combines offline initialization, learning from prior physical feedback, and final policy refinement (Appendix A.4). To assess the *online learning design*, a controlled study starts all learning methods from the same offline Raw-FM + J policy, with matched physical-evaluation caps and update budgets. Figure 3 tracks discoveries against each branch's growing archive; Table 3 pairs these discoveries with the resulting policies' ordinary generation beyond $\mathcal{D}_0$.

Data and physical evaluation. We study stellarator-symmetric, two-field-period boundaries with 168 free Fourier coefficients. The initial corpus $\mathcal{D}_0$ contains 30,598 records from finite-pressure equilibrium preparation and automated boundary optimization; J-direct selects 3,936 entries (Appendix A.2). DESC (Panici et al., 2023) solves fixed-boundary MHD equilibria under prescribed pressure, current, and flux settings. Request validity ($V = 1$) requires geometric admissibility, recovered MHD equilibrium, aspect-ratio error at most 5%, and rotational-transform error at most 0.05. Joint success ($G_\tau = 1$) further requires normalized QS residual below $\tau = 10^{-3}$ and nonnegative off-axis magnetic well. Strict success tightens QS to 5×10^{-4}. Success rates include all designated outputs; QS and well summaries use recovered, measurable equilibria.

Offline baselines. **DFF** uses the author's PCA–DDPM implementation (Padidar et al., 2025), retrained on $\mathcal{D}_0$. **DFF-PCA-Joint** further trains this model on J-direct data while retaining its full-data PCA (Appendix A.6). **Latent-FM** and **Raw-FM** compare a learned 24-dimensional codec with full-boundary generation under matched flow architectures and training updates. **Raw-FM + J** shares Raw-FM's pretrained weights and continuation budget, replacing full-data continuation with J-direct (Appendix A.7).

Ordinary evaluation. The primary *target-domain* panel contains 32 requests from the development and refinement domain, including the eight used for online feedback. We additionally assess *request transfer* on 64 additional requests, keeping the policies fixed without further adaptation. Each model supplies 16 preselected ordinary outputs per request: 512 and 1,024 evaluations, respectively. Models share requests, designated positions, physics, and acceptance criteria. Qualified novel yield is the fraction of outputs satisfying $G_\tau = 1$ and novel relative to $\mathcal{D}_0$ (Section 3.3). Distinct designs

Table 1: **End-to-end generation quality.** QS: median [90th percentile], $\times 10^{-4}$. W^+: fraction with nonnegative well throughout; f_W^+: mean nonnegative-node fraction. QS/well use recovered, measurable designs. Recovery, Joint (G_τ), and Strict include all outputs. Rates are percentages; best per panel is bold.

Model	Recovery ↑	QS ↓	W^+ ↑	f_W^+ ↑	Joint ↑	Strict ↑
Target-domain requests						
DFF	79.88	15.80 [31.76]	62.59	84.38	8.79	0.39
DFF-PCA-Joint	72.46	12.69 [24.55]	64.42	86.93	16.60	1.37
Latent-FM	74.02	17.40 [39.19]	65.70	86.68	10.55	3.12
Raw-FM	75.39	8.89 [33.47]	64.77	85.93	25.78	16.02
Raw-FM + J	87.30	**3.80** [11.88]	74.05	89.58	55.66	41.99
SCOPE	**90.04**	4.32 [**8.98**]	**85.25**	**94.84**	**71.29**	**49.22**
Request transfer						
DFF	83.59	14.25 [27.74]	66.36	87.04	10.45	0.29
DFF-PCA-Joint	75.10	11.62 [20.83]	68.01	89.45	17.19	1.37
Latent-FM	76.76	13.47 [33.59]	67.43	89.54	19.53	6.25
Raw-FM	80.86	7.27 [33.74]	64.86	86.46	30.37	19.53
Raw-FM + J	**91.31**	3.36 [11.75]	73.05	90.56	60.84	48.44
SCOPE	91.02	**3.12** [**10.21**]	**77.58**	**93.01**	**65.33**	**53.32**

additionally require pairwise separation under the same three distances. Counting rules are fixed across models; request selection is detailed in Appendix C.3.

4.2 END-TO-END PHYSICAL QUALITY

Full boundaries and joint-quality initialization. Table 1 shows the value of preserving the physical design coordinates. Raw-FM reaches 25.78% joint success in the target domain, compared with 10.55% for Latent-FM. J-direct raises Raw-FM's success to 55.66%, and from 30.37% to 60.84% on the transfer panel. These two continuations share their initial weights and update budget, identifying the benefit of joint-quality training data. DFF-PCA-Joint also improves on DFF, reaching 16.60% and 17.19% joint success. The full-boundary policy supplies a stronger foundation for learning from physical feedback.

Quality of the complete system. The final SCOPE generator achieves 71.29% and 65.33% joint success, respectively: $8.1\times$ and $6.3\times$ DFF's rates, and $4.3\times$ and $3.8\times$ DFF-PCA-Joint's. Its median QS residuals are 4.32×10^{-4} and 3.12×10^{-4}; nonnegative well throughout the evaluated profile occurs in 85.25% and 77.58% of recovered designs. Relative to Raw-FM + J, this generator improves the QS upper tail and raises the mean nonnegative-well fraction from 89.58% to 94.84% and from 90.56% to 93.01%, respectively. The gains extend to the stricter QS gate. Joint acceptance and continuous physical measurements show that the complete pipeline supplies promising candidates across both request panels (Appendix C.1).

4.3 REUSABLE COVERAGE AND REQUEST TRANSFER

Table 2 evaluates each fully trained generator's qualified repertoire beyond $\mathcal{D}_0$, in the target domain and under request transfer. All counted designs satisfy the same joint-quality gate.

A broader repertoire of qualified designs. In the target domain, SCOPE yields 34.38% qualified novel outputs and 102 distinct designs, compared with 8.40% and 42 for DFF, and 11.33% and 51 for DFF-PCA-Joint. It thus supplies twice as many distinct qualified alternatives as the quality-refined DFF baseline. The improvement is directly available to a design sweep through ordinary generation.

The offline comparisons separate quality from breadth. Latent-FM, Raw-FM, and Raw-FM + J increase joint success from 10.55% to 25.78% to 55.66%, with qualified novel coverage of 23, 26, and 29 designs. Full-boundary modeling and quality initialization establish a useful physical prior; DFF's 42 designs illustrate the value of broad offline coverage. SCOPE combines strong joint quality

Table 2: **End-to-end qualified coverage beyond $\mathcal{D}_0$.** Joint and qualified novel yields are percentages of ordinary outputs. Distinct counts are separated qualified novel designs. Requests and allocation match Table 1; higher is better, and best per panel is bold.

	Target-domain requests			Request transfer		
Model	Joint (%)	Novel (%)	Distinct	Joint (%)	Novel (%)	Distinct
DFF	8.79	8.40	42	10.45	9.96	97
DFF-PCA-Joint	16.60	11.33	51	17.19	**12.01**	**109**
Latent-FM	10.55	5.47	23	19.53	8.50	72
Raw-FM	25.78	5.08	26	30.37	6.25	59
Raw-FM + J	55.66	9.96	29	60.84	9.67	68
SCOPE	**71.29**	**34.38**	**102**	**65.33**	11.52	88

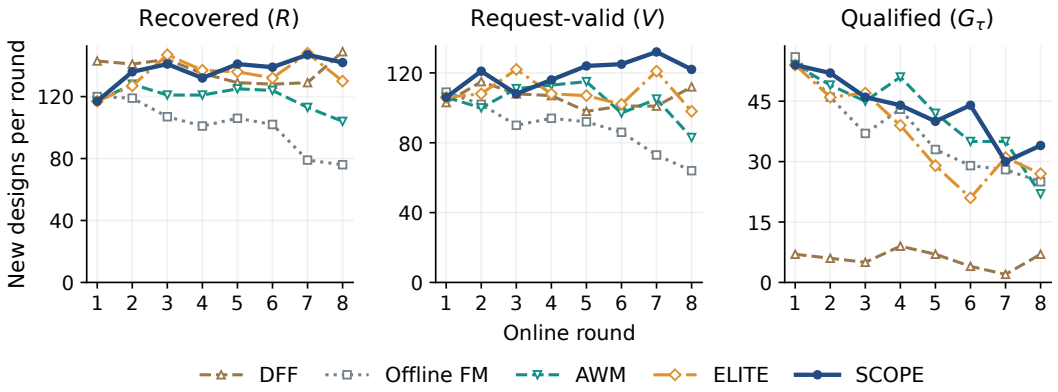


Figure 3: **Discovery continues as the archive grows.** New recovered representatives and their request-valid and qualified subsets in each round. Each branch starts from $\mathcal{D}_0$ and expands its own certified archive under the same evaluation cap.

with a larger repertoire by learning from physical feedback beyond the original corpus. Once trained, SCOPE averages 60.0 distinct designs at half the ordinary evaluation budget, exceeding both DFF baselines at their full budgets (Appendix C.2).

Reusing the generator across requests. On additional requests, SCOPE retains the highest joint success at 65.33% and supplies 88 distinct qualified novel designs. Its 11.52% qualified novel yield exceeds all three offline FM models, whose coverage counts are 72, 59, and 68. Coverage transfers less fully than quality: DFF and DFF-PCA-Joint supply 97 and 109 distinct designs, respectively. SCOPE thus expands the qualified repertoire beyond the offline FM baselines on both panels, while the DFF baselines retain broader coverage on the transfer panel. Policy updates compare responses within each queried condition and replay designs at the requests they fulfill. Expanding coverage across the condition space is a further learning objective.

4.4 CONTINUED DISCOVERY AGAINST A GROWING ARCHIVE

To compare online learning rules, we run independent eight-round searches with fixed DFF, **Offline FM**, **AWM** (Xue et al., 2026), **ELITE**, and SCOPE. Offline FM holds Raw-FM + J fixed; the learning methods start from these weights and collect their own feedback. AWM applies advantage-weighted matching to current feedback; ELITE is SCOPE's positive replay ablation with the signed term disabled. Each round uses the same candidate budget and at most 256 physical evaluations. Archives start at $\mathcal{D}_0$ and admit novel, separated geometries passing R. Their valid and qualified subsets measure task-relevant discovery (Appendix A.5).

Table 3: **Search yield and final coverage, with SCOPE ablations.** Online counts accumulate eight rounds; final ordinary V/G_τ designs are independently separated beyond $\mathcal{D}_0$.

	Online discoveries		Final generator: online requests			
Policy	V	G_τ	Valid (%)	Joint (%)	Novel V	Novel G_τ
Offline FM	710	297	95.70	58.59	28	12
AWM	830	333	96.48	57.81	53	23
ELITE	872	294	96.09	51.56	90	36
SCOPE-Uniform	890	352	**98.44**	**59.38**	86	43
SCOPE-Current	**985**	**365**	97.66	49.61	108	41
SCOPE	954	344	97.27	50.00	**120**	**51**

Useful discoveries continue as knowledge grows. Over eight rounds, DFF and SCOPE discover 1,098 and 1,095 recovered representatives (Figure 3). SCOPE supplies 954 request-valid and 344 qualified designs, compared with 845 and 47 for DFF and 710 and 297 for Offline FM. Its final round adds 122 valid and 34 qualified designs. The lead over ELITE grows in the last four rounds, reaching 82 valid and 50 qualified discoveries (9.4% and 17.0%). A second online seed reproduces this advantage (Appendix C.6).

AWM discovers 830 valid and 333 qualified designs. SCOPE gains 14.9% in valid discovery, chiefly in the last four rounds (503 versus 400); qualified totals are similar.

4.5 FROM EXPLORATION TO REUSABLE GENERATION

Table 3 compares cumulative discoveries and final ordinary coverage on the eight online requests. Learning methods share the offline start, acquisition rules, rewards, and update budgets. We ablate two replay choices within SCOPE: *uniform replay* samples valid designs equally within requests; *current-round priority* boosts discoveries only in their acquisition round. Both retain cumulative valid replay and signed updates. Transfer appears in Appendix C.5.

Turning feedback into a broader repertoire. AWM expands ordinary valid and qualified coverage from the offline policy's 28 and 12 designs to 53 and 23. SCOPE supplies 120 and 51 ($2.3\times$ and $2.2\times$ AWM), at joint success of 50.00% versus 57.81%.

Signed feedback improves on positive replay. ELITE generates 90 valid and 36 qualified distinct designs. Signed feedback raises these counts by 33.3% and 41.7%, at similar joint success (51.56% and 50.00%), improving both search yield and ordinary coverage. A matched-feedback diagnostic shows generation favoring higher-advantage neighborhoods (Appendix C.4).

Replay priority shapes discovery and generation. Within SCOPE, discovery weighting expands ordinary valid and qualified coverage from 86 and 43 under uniform replay to 120 and 51. Uniform replay yields higher joint success (59.38%) and more qualified discoveries (352 versus 344); sustained weighting yields more valid discoveries (954 versus 890).

Restricting priority to the acquisition round yields 985 valid and 365 qualified discoveries. Sustaining priority across rounds expands final coverage from 108 valid and 41 qualified designs to 120 and 51, at similar joint success (49.61% and 50.00%). These ablations show how replay within SCOPE shifts the balance between search yield and reusable generation (second seed in Appendix C.6).

5 CONCLUSION

SCOPE expands the repertoire of request-valid MHD equilibria, guided by QS and magnetic-well quality. In the target domain, ordinary generation supplies more distinct qualified candidates, and the closed loop sustains valid and qualified discovery against a growing archive. Request transfer keeps strong physical quality, with broader cross-condition coverage remains for future development.

AI USE STATEMENT

Generative AI assistants supported language polishing, and assisted software implementation and debugging, analysis scripting, manuscript revision, and figure prototyping. The authors directed the research and take responsibility for the scientific claims, code, analyses, and final manuscript.

REFERENCES

David Bindel, Matt Landreman, and Misha Padidar. Understanding trade-offs in stellarator design with multi-objective optimization. *Journal of Plasma Physics*, 89(5):905890503, 2023. doi: 10.1017/S0022377823000788.

David Brookes, Hahnbeom Park, and Jennifer Listgarten. Conditioning by adaptive sampling for robust design. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 773–782. PMLR, 2019. URL https://proceedings.mlr.press/v97/brookes19a.html.

Santiago Cadena, Andrea Merlo, Emanuel Laude, Alexander Bauer, Atul Agrawal, Maria Pascu, Marija Savtchouk, Lukas Bonauer, Enrico Guiraud, Stuart Hudson, and Markus Kaiser. ConStellaration: A dataset of QI-like stellarator plasma boundaries and optimization benchmarks. In *Advances in Neural Information Processing Systems*, volume 38. Curran Associates, Inc., 2025. doi: 10.52202/085713-0520.

Maxence Faldor, Félix Chalumeau, Manon Flageat, and Antoine Cully. MAP-Elites with descriptor-conditioned gradients and archive distillation into a single policy. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 138–146. Association for Computing Machinery, 2023. doi: 10.1145/3583131.3590503.

Adam Gaier, Alexander Asteroth, and Jean-Baptiste Mouret. Discovering representations for black-box optimization. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pp. 103–111. Association for Computing Machinery, 2020. doi: 10.1145/3377930.3390221.

Andrew Giuliani, Florian Wechsung, Antoine Cerfon, Georg Stadler, and Matt Landreman. Single-stage gradient-based stellarator coil design: Optimization for near-axis quasi-symmetry. *Journal of Computational Physics*, 459:111147, 2022. doi: 10.1016/j.jcp.2022.111147.

Andrew Giuliani, Eduardo Rodríguez, and Marina Spivak. A comprehensive exploration of quasisymmetric stellarators and their coil sets. *Journal of Plasma Physics*, 91(5):E128, 2025. doi: 10.1017/S0022377825000509.

Shashank Hegde, Sumeet Batra, K. R. Zentner, and Gaurav Sukhatme. Generating behaviorally diverse policies with latent diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 7541–7554. Curran Associates, Inc., 2023. doi: 10.52202/075280-0332.

Matt Landreman and Elizabeth Paul. Magnetic fields with precise quasisymmetry for plasma confinement. *Physical Review Letters*, 128(3):035001, 2022. doi: 10.1103/PhysRevLett.128.035001.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Misha Padidar, Teresa Huang, Andrew Giuliani, and Marina Spivak. Diffusion for fusion: Designing stellarators with generative AI. In *NeurIPS 2025 Workshop on AI for Science*, 2025. URL https://arxiv.org/abs/2511.20445.

Dario Panici, Rory Conlin, Daniel W. Dudt, Kaya Unalmis, and Egemen Kolemen. The DESC stellarator code suite. part 1. quick and accurate equilibria computations. *Journal of Plasma Physics*, 89(3):955890303, 2023. doi: 10.1017/S0022377823000272.

Shuchen Xue, Chongjian Ge, Shilong Zhang, Yichen Li, and Zhi-Ming Ma. Advantage weighted matching: Aligning RL with pretraining in diffusion models. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=nLY2pOYBrJ.

A PHYSICAL TASK AND IMPLEMENTATION

A.1 BOUNDARY REPRESENTATION AND CONDITIONS

We use stellarator-symmetric, two-field-period boundaries with poloidal and toroidal Fourier orders up to six, in the real product basis used by DESC. There are 169 retained coefficients before fixing $R_{0,0} = 1$, leaving 168 free coordinates. Training standardization is fitted on the training partition and fixed throughout feedback refinement.

The model receives aspect ratio and signed vacuum rotational-transform targets. The latter is the mean of DESC's `iota vacuum` component at ten radial nodes from $\rho = 0.1$ to 0.95, computed on the recovered finite-pressure equilibrium. Each request binds c to the profiles, current, flux, and dimensional scale η of its source equilibrium. These prescribed quantities remain fixed while the generated boundary is recovered. Training examples retain their native conditions; accepted feedback retains its generating request.

A.2 CORPUS CONSTRUCTION AND PHYSICAL SETTINGS

The corpus combines finite-pressure equilibrium preparation from QUASR-derived boundaries with records from an automated physical-optimization workflow. The preparation campaign targets $\beta \simeq 2\%$ through kinetic profiles and magnetic normalization. Subsequent optimization includes a campaign at major-radius scale $1.5\,\mathrm{m}$ with fixed density and temperature profiles: on-axis electron density $2 \times 10^{19}\,\mathrm{m}^{-3}$, electron temperature $2\,\mathrm{keV}$, and ion temperature $0.4\,\mathrm{keV}$. In this campaign, β is an equilibrium outcome under the prescribed profiles. The stored physical settings are retained per record and per evaluation request.

The corpus contains 33,924 native records and 33,810 normalized geometry identities, partitioned into 30,598 training and 3,326 validation records. Identical normalized geometries are grouped before a fixed hash split, giving zero exact-geometry overlap between partitions. Records receive harmonized native aspect, signed vacuum-transform, QS, and well annotations before quality selection. J-direct contains 3,936 entries and 3,917 distinct geometries. All compared generators receive the same full offline corpus before their method-specific refinement.

J-direct computes the fractional midrank of each example's QS violation among its K nearest neighbors in standardized request space. The selected set $\mathcal{D}_J$ retains examples whose local rank is at most q and whose magnetic well is nonnegative throughout the evaluated interior. Table 4 gives K and q.

A.3 JOINT ACCEPTANCE AND NUMERICAL EVALUATION

We compute fixed-boundary MHD equilibria with DESC, holding each generated boundary and its prescribed physical settings fixed. All generators use the same force-balance objective and at most 20 cumulative solver iterations. Physical qualification is determined from the resulting equilibrium using the checks below.

Physical recovery $R = 1$ requires geometric admissibility, normalized force-balance RMS at most 10^{-4} on both the working and denser validation grids, positive Jacobian, and nested surfaces. Request validity $V = 1$ additionally requires $|\hat{A} - A^\star|/|A^\star| \le 0.05$ and $|\hat{\iota} - \iota^\star| \le 0.05$. The joint-quality gate for evaluation is

$$G_\tau(x;u) = 1 \quad \Longleftrightarrow \quad V(x;u) = 1, \quad Q < \tau, \quad W(\rho_j) \ge 0 \text{ at all 50 off-axis nodes.} \tag{7}$$

The primary threshold is $\tau = 10^{-3}$; the stricter analysis uses 5×10^{-4}. The QS score is the RMS of normalized symmetry-breaking Boozer coefficients over ten radial surfaces, with a de-aliased angular projection and the prescribed reference magnetic-field scale. These measurements establish task validity and then assess QS and magnetic-well quality for early-stage design.

Table 4: **Full-boundary policy configuration.** The boundary representation and sampler remain fixed during online learning.

Setting	Value
Velocity network	20 residual blocks; width 1,024; request width 512
Boundary / request dimensions	168 / 2
J-direct selection	$K = 256$ neighbors; local QS rank $q = 0.25$
Sampler	32 steps; Heun with final Euler step; 63 evaluations
Guidance strength	1 (conditional velocity)
Precision / gradient clipping	BF16 forward, FP32 error; gradient norm cap 1

A.4 Archive-guided policy refinement implementation

Physical reward. For a request-valid, measurable endpoint, the physical utility in Equation 2 is

$$U(x) = \frac{1}{2}\min\left(1, \frac{10^{-3}}{\max(Q(x), 10^{-12})}\right) + \frac{1}{2m}\sum_{j=1}^{m}\mathbb{I}[W(x,\rho_j) > 0], \qquad \rho_j > 0. \tag{8}$$

The utility is zero when quality measurements are unavailable. Online utility uses positive off-axis well nodes; joint acceptance uses nonnegative well throughout. The reward coefficients are $\lambda_N = 1$, $\lambda_d = 0.15$, $\lambda_Q = 0.05$, and $\lambda_{\text{fail}} = 1$. Advantages are normalized within each request's evaluated responses with $\sigma_{\min} = 0.1$ and clipped at $A_{\max} = 2$. A batch standard deviation below 10^{-4} gives zero advantages for that request.

Replay and signed updates. Positive replay balances requests and weights online endpoints by $1 + 4m_{ki}$, retaining their acquisition-time discovery status. We normalize endpoint weights within requests and assign equal mass to requests. The quality pool μ_{quality} samples uniformly within its J-direct and previously qualified subsets, with batch allocations specified below. A separate batch of 128 current physical responses, including unsuccessful outcomes, supplies the signed objective and reference penalty. Both use fresh Gaussian noise and $t \sim \mathcal{U}(0.001, 0.999)$. Matching errors are evaluated in FP32 after BF16 network evaluations.

The old-error scale floor is $s_{\min} = 0.1$ and surrogate clipping uses $\kappa = 0.2$. At the first minibatch of each window, we set $\lambda_A = \min(10, 0.5\|g_{\text{pos}}\|_2 / \max(\|g_{\text{fb}}\|_2, 10^{-12}))$ using the two parameter-gradient norms, then hold it fixed. The reference penalty is

$$\mathcal{R}(\theta, \theta_{\text{ref}}) = \frac{\mathbb{E}\|v_\theta(x_t, t, c) - v_{\theta_{\text{ref}}}(x_t, t, c)\|_2^2/d}{\max(\mathbb{E}\|x - \epsilon\|_2^2/d, 0.1)}, \tag{9}$$

with $\lambda_{\text{ref}} = 0.1$. The reference is the initial online policy; the old policy in Equation 5 is refreshed each round.

AdamW uses learning rate 10^{-6}, $(\beta_1, \beta_2) = (0.9, 0.999)$, weight decay 0.01, and gradient-norm clipping at one. Each round permits 256 updates. Every 16 updates, normalized velocity drift is measured from the round's old policy on all current feedback endpoints, with fixed Gaussian noise and evenly spaced times from 0.005 to 0.995. If drift exceeds 0.02, we return the previous accepted iterate. The reported runs complete all eight windows of 256 updates; evaluation uses their terminal accepted raw weights.

Full-system training. Tables 1 and 2 evaluate the final generator after iterative physical-feedback training, which alternates candidate generation, physical evaluation, and learning from the resulting feedback. Qualified designs collected during this process remain in replay throughout the final eight-round refinement. Each positive batch contains 128 J-direct examples, 64 qualified designs from preceding searches, and 64 valid designs accumulated during this refinement.

The final eight-round refinement on the eight target-domain feedback requests uses 2,048 physical-evaluation positions and 2,048 update steps. Each request receives 1,024 proposals per round, with 16 preselected ordinary and 16 geometrically selected evaluations. Selection prioritizes separated novel admissible candidates, then proximity to the frontier. The reference contains the original corpus

and admissible proposal history, including midpoint-evaluation geometries; midpoint physical labels remain reserved for evaluation.

A.5 CONTROLLED CLOSED-LOOP STUDY

Figure 3 compares five branches of an independent eight-round experiment; Table 3 also includes two replay ablations of SCOPE. Every archive starts from $\mathcal{D}_0$: 30,598 records and 30,492 unique geometries. DFF and Offline FM remain fixed. All learning variants share the offline Raw-FM + J weights in Appendix A.7 and collect their own feedback under the same requests, acquisition rules, reward, and update budget. ELITE ablates SCOPE's signed term, retaining request-balanced valid replay, discovery weights, quality anchors, reference anchoring, and drift checks.

Archive admission and learning. Each round freezes its reference during selection and evaluation. Novel candidates passing R are ordered by geometry identity and greedily separated under all three distances before admission. Each member retains its recovered equilibrium and physical record. The plotted V and G_τ curves count subsets of this admitted representative set. Positive replay uses $V = 1$. ELITE and SCOPE use 192 J-direct anchors and 64 accumulated valid endpoints per batch; their fixed velocity reference is the common offline starting policy. All learning methods complete 256 accepted updates per round. Archives grow from their own branch's rollouts.

Online allocation. Each of eight requests receives 1,024 candidates, 16 preselected ordinary evaluations, and up to 16 separated novel frontier evaluations per round. Unfilled frontier slots remain unused; physical failures consume their positions. Each branch generates 65,536 candidates and has at most 2,048 physical positions over eight rounds. Actual positions are 2,048 for DFF, 1,919 for Offline FM, 1,953 for AWM, 2,042 for ELITE, and 2,023 for SCOPE; every branch completes all 1,024 ordinary positions. Reused physical results consume the same evaluation budget. Appendix C.6 reports a second online seed for DFF, ELITE, and SCOPE.

Ordinary generation after exploration. After the eighth update, each final policy is held fixed and evaluated on 32 preselected ordinary samples per request. Table 3 uses the eight online requests, totaling 256 outputs per policy. Table 8 additionally evaluates Offline FM and the replay variants on eight further requests with the same allocation. These are distinct from the 64-request transfer panel in Tables 1 and 2. Within each panel, methods share designated requests and sample positions. Each qualification set is independently separated against $\mathcal{D}_0$. Cumulative online discoveries summarize the policies used to collect feedback; ordinary generation measures the policy after the final update.

Advantage Weighted Matching. We adapt the authors' AWM implementation[1] to the common physical-feedback protocol. Updates use current-round endpoints with the clipped advantage-weighted matching objective, error ratio $\rho = \exp(e_{\bar{\theta}_k} - e_\theta)$, clipping range $[0.8, 1.2]$, and unit-weight velocity-reference penalty. The rollout policy remains frozen during each round's updates; the reference is the common offline policy. Each batch samples 64 endpoints and averages six independent draws of Gaussian noise and uniform time per endpoint. AWM uses the same AdamW settings, 256 updates per round, and drift checks described above.

SCOPE with uniform replay. This ablation replaces the full configuration's online replay weights $1 + 4m_{ki}$ with one. It retains the cumulative valid pool, equal request mass, signed objective, quality anchors, reference anchoring, and update budget. Both runs share their offline initialization and first pre-update rollout; subsequent feedback comes from each policy's own search. The variant with uniform replay uses 1,961 online physical positions and completes all 2,048 updates. Its final evaluation shares the full configuration's ordinary request/noise positions and the panel split above.

SCOPE with current-round priority. This ablation applies extra discovery weight only during the round in which a representative is acquired, while retaining the same cumulative valid replay pool and learning objective. Older valid examples remain in replay with unit weight. It uses 2,038 online physical positions and completes all 2,048 updates. Together with the full configuration, these ablations compare uniform allocation, current-round priority, and persistent discovery priority within SCOPE (Table 3).

[1] https://github.com/scxue/advantage_weighted_matching

A.6 DFF ON THE SHARED OFFLINE CORPUS

We retrain the author's PCA–DDPM implementation (Padidar et al., 2025) on all 30,598 unfiltered training entries, using the existing validation split. An exact basis adapter maps the native boundaries into the author's surface representation; PCA with 50 components is fitted on training data only. We retain the author's normalization, residual denoising network, noise-prediction objective, and 200-step ancestral sampler. The condition inputs contain aspect ratio and signed vacuum rotational transform, together with the fixed field-period and helicity values.

Training uses AdamW with learning rate 5×10^{-4} and batch size 4,096. A predefined validation-loss rule checks progress every 250 epochs, continuing when the recent loss improves by more than 1%, up to 1,000 epochs. Training stops at epoch 750 after 5,250 updates; checkpoint selection uses fixed-noise validation denoising loss, without consulting the physical evaluation results. Both the published PCA dimension and all stopping choices are fixed before evaluation.

The baseline uses the same requests, preselected ordinary positions, DESC equilibrium evaluation, QS definition, and magnetic-well criteria as the flow models. Failed samples remain in the success-rate denominator.

DFF-PCA-Joint continues this checkpoint for 1,750 updates on the 3,936 J-direct entries, keeping the full-data PCA, coordinate and condition normalization, network, and sampler fixed. It uses a fresh AdamW optimizer at learning rate 5×10^{-4} and replacement batches of 4,096, for 7,000 total updates including pretraining. Evaluation uses the terminal checkpoint on both ordinary panels.

A.7 MATCHED FLOW REPRESENTATION AND INITIALIZATION BASELINES

The representation study compares full-boundary flow matching with a learned 24-dimensional latent representation on the same 30,598 training entries and existing validation split. Both flow backbones use 20 residual blocks of width 1,024 and condition-token width 512, with matched hidden-layer initialization. Only the input/output dimensions differ: 168 free boundary coordinates for Raw-FM and 24 for Latent-FM. They contain 90.81M and 90.52M flow parameters, respectively. A separately trained 48.19M-parameter codec supplies the latent representation. It uses frequency-structured decoder heads, a high-frequency Gaussian reconstruction term, and a KL regularizer. The decoder receives the requested condition and latent vector, without an encoder skip connection or source boundary at generation time.

The codec is trained for 10,000 updates on training entries only and then frozen; latent normalization is fitted on the same training partition. Both flow models use two stages of 10,000 updates, batch size 512, AdamW with learning rate 10^{-4} decaying to 10^{-6}, and moving-average decay 0.999. Each continuation starts from its preceding moving average with a fresh optimizer. Raw-FM + J shares the first Raw-FM stage and replaces only the second stage's full-data examples with the 3,936 J-direct entries. Consequently, the two Raw-FM continuations have identical initialization and update budgets. All final models use the fixed terminal moving average, without selecting checkpoints by physical evaluation.

Sampling uses 32 integration steps (Heun with a final Euler step, 63 velocity evaluations), guidance scale one, and the same ordinary request/position allocation as the other models. The latent decoder adds a separately accounted representation cost. Every generated boundary is physically evaluated without replacing failures. Full-data, J-direct, and feedback models all use the complete original training set as the novelty reference.

B GEOMETRIC NOVELTY AND COUNTING

Let $b(x)$ denote the complete boundary vector after restoring the fixed major-radius coefficient. With a fixed surface-normal factor F, low-frequency projection P_{low}, dense surface basis V_{surf}, and

reference minor radius a, the three distances are

$$d_1(x,z) = \|(b(x) - b(z))F\|_2, \tag{10}$$

$$d_2(x,z) = \|P_{\text{low}}(b(x) - b(z))F\|_2, \tag{11}$$

$$d_3(x,z) = \frac{\text{RMS}(V_{\text{surf}}[b(x) - b(z)])}{a}. \tag{12}$$

The low-frequency mask retains $|m| + |n| < 9$. Dense geometry is sampled on a 32×24 angular grid over one field period. Calibration uses 1,024 original-data geometries nearest to a fixed reference condition in standardized request space. We pair each geometry with its leave-one-out nearest neighbor under the full normal metric, then take the 95th percentile of each of the three distances on these pairs. This fixes $(\delta_1, \delta_2, \delta_3) \simeq (0.00449714, 0.00449710, 0.03771964)$ before feedback learning. The normal threshold is about 0.45% of the major-radius normalization; the dense-displacement threshold is 3.77% of the reference minor radius. The scales thus compare generated shapes with characteristic geometric spacing in the original corpus. During novelty evaluation, each metric independently searches the full reference for its nearest neighbor; all three distances must strictly exceed their fixed scales.

For distinct-design counts, eligible novel candidates are ordered by geometry identity. A candidate is retained if its distance from every retained representative exceeds the scale under every metric. The rule applies across requests. Figure 3 first separates recovered candidates and then counts their V/G_τ subsets. The online columns of Table 3 accumulate these subsets; its final-generator columns independently separate each eligibility set. The fixed-reference ordinary comparisons search all 30,492 unique training geometries.

C Additional results and evaluation details

C.1 Independent physical quality

Ordinary sample allocations are specified in Section 4.1; recovered counts for the compared models appear in Table 5. QS and well statistics use the recovered, physically measurable population without conditioning on the request gate, the QS threshold, or the well gate. Recovery, joint success, and strict success retain all designated slots in their denominators.

Our QS residual is the RMS of the symmetry-breaking Boozer coefficients over ten radial surfaces, normalized by the reference field B_{ref}. Diffusion for Fusion (Padidar et al., 2025) reports a magnetic-surface-area-weighted relative field-energy error. Both quantify departure from quasisymmetry, but the normalizations and averaging measures differ; our numerical QS values use the single fixed definition throughout all compared models.

For a recovered design with m evaluated off-axis radial nodes, define $f_W^+(x) = m^{-1} \sum_{j=1}^{m} \mathbb{I}[W(x, \rho_j) \geq 0]$. The design-level W^+ fraction in Table 1 is the proportion with $f_W^+ = 1$; the adjacent column reports the mean of f_W^+. The on-axis zero is excluded from both. The radial well average $\bar{W} = \int_0^1 W(\rho)\, d\rho$ uses trapezoidal quadrature on the full saved profile. Figure 4 and Table 5 give distributions and request errors.

Request fidelity and continuous quality. On target-domain requests, Raw-FM reduces median aspect-ratio error from Latent-FM's 2.428% to 0.682%, and transform error from 0.02363 to 0.00907 (Table 5). J-direct brings these errors to 0.402% and 0.00493; SCOPE maintains comparably accurate request matching at 0.397% and 0.00578. Figure 4 complements these measurements: quality initialization places much of the recovered population at low QS residuals, and SCOPE improves the QS upper tail and the fraction of designs with nonnegative well throughout. The continuous measurements connect joint success to request fidelity and complementary physical quality.

From task validity to joint qualification. Table 6 traces ordinary outputs through MHD recovery, request matching, QS, and joint qualification, with all rates measured against the full sample allocation.

On target-domain requests, Raw-FM raises validity from Latent-FM's 51.76% to 67.77%; J-direct raises it to 82.81%. DFF and DFF-PCA-Joint both reach 71.09% validity, with joint success of 8.79%

Table 5: **Recovered equilibria and request errors.** Errors use recovered equilibria; aspect-ratio error is relative.

Model	Recovered	Median e_A (%)	Median e_ι
Target-domain requests			
DFF	409	1.119	0.01371
DFF-PCA-Joint	371	0.859	0.01190
Latent-FM	379	2.428	0.02363
Raw-FM	386	0.682	0.00907
Raw-FM + J	447	0.402	0.00493
SCOPE	461	0.397	0.00578
Request transfer			
DFF	856	1.170	0.01251
DFF-PCA-Joint	769	0.829	0.01095
Latent-FM	786	2.021	0.02045
Raw-FM	828	0.623	0.00750
Raw-FM + J	935	0.363	0.00411
SCOPE	932	0.340	0.00401

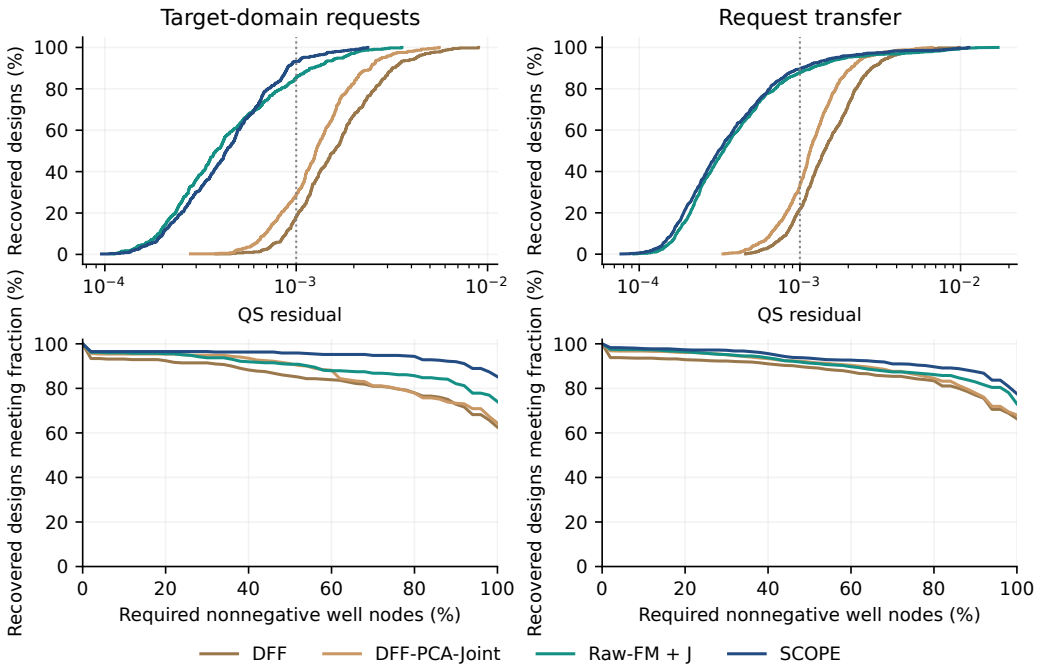


Figure 4: **Physical quality beyond a binary success gate.** DFF, DFF-PCA-Joint, Raw-FM + J, and SCOPE distributions include all recovered, physically measurable outputs. Top: empirical cumulative QS distributions; the dotted line marks the main QS threshold. Bottom: fraction attaining each required proportion of nonnegative radial well nodes. The endpoint at 100% is the design-level W^+ fraction.

and 16.60%, respectively. SCOPE reaches 86.52% validity; 82.23% of all outputs also pass QS, and 71.29% meet the joint criterion. On the transfer panel, its valid and joint rates are 88.38% and 65.33%, compared with 87.89% and 60.84% for Raw-FM + J. The progression connects accurate conditional generation with increasingly demanding physical quality.

Table 6: **From recovered equilibria to qualified designs.** Cumulative pass rates (%) over all ordinary outputs. R: admissible MHD equilibria; V: additional request matching; V + QS: additional $Q < 10^{-3}$; G_τ: additional nonnegative well throughout. Best per panel is bold.

Model	Recovered R	Valid V	V + QS	Qualified G_τ
Target-domain requests				
DFF	79.88	71.09	13.28	8.79
DFF-PCA-Joint	72.46	71.09	20.90	16.60
Latent-FM	74.02	51.76	12.89	10.55
Raw-FM	75.39	67.77	38.28	25.78
Raw-FM + J	87.30	82.81	73.63	55.66
SCOPE	**90.04**	**86.52**	**82.23**	**71.29**
Request transfer				
DFF	83.59	77.83	17.77	10.45
DFF-PCA-Joint	75.10	73.44	24.80	17.19
Latent-FM	76.76	54.79	25.29	19.53
Raw-FM	80.86	73.34	47.07	30.37
Raw-FM + J	**91.31**	87.89	79.98	60.84
SCOPE	91.02	**88.38**	**81.15**	**65.33**

C.2 QUALIFIED COVERAGE AND SWEEP BUDGETS

Evaluation budget and design yield. Each request's 16 ordinary positions are randomly ordered with seed 20260916, using the same orders across models. At allocation $k/16$, we retain the first k positions for every request. We repeat this construction 128 times and recompute separated representatives in every subset. Lines in Figure 5 show mean counts; bands show the 10th–90th percentiles across subsets. Full-budget points reproduce the two panels of Table 2. The horizontal axis counts ordinary evaluations of each fixed final generator.

On target-domain requests, SCOPE yields an average of 60.0 distinct qualified designs at half budget, exceeding DFF's 42 and DFF-PCA-Joint's 51 at full budget. The target-domain curve connects the broader learned repertoire to a practical sweep benefit: fewer physical evaluations are needed to obtain a substantial set of distinct qualified alternatives. The transfer curve shows the DFF baselines' broader coverage at larger budgets, complementing the quality gains in Table 1.

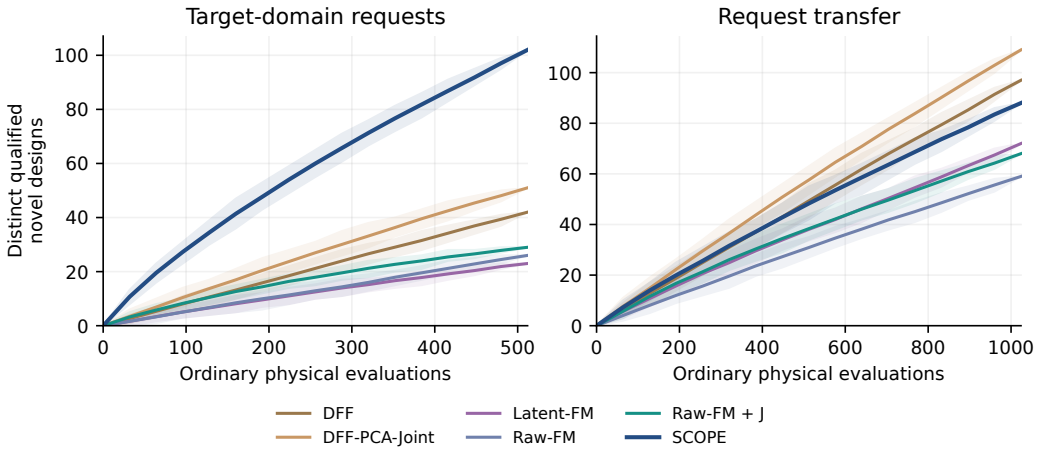


Figure 5: **Qualified design yield across evaluation budgets.** Distinct qualified designs beyond $\mathcal{D}_0$ versus ordinary physical evaluations. Lines average request-balanced subsets; bands show their 10th–90th percentiles.

C.3 Evaluation request panels

The 32-request target-domain panel is the development and refinement panel; eight of its requests supply the online feedback. The request-transfer panel contains 64 additional requests selected without physical-quality labels, excluding training geometry identities, previously used requests, and reserved evaluation entries. Both panels draw source geometries from the generator's validation partition. Each request combines its condition targets with prescribed physical settings. Transfer labels are reserved for evaluation, and each policy is used without further adaptation. The split measures reuse across requests within this design setting. All models use the DESC evaluation settings in Appendix A.3.

C.4 Effect of signed feedback on local generation

We compare the common offline policy with ELITE and SCOPE after their first 256 training steps on the same 250 physically evaluated responses. Each policy generates 8,192 ordinary samples for each of the eight online requests, using shared fresh Gaussian inputs. Within each request, we assign each sample to the nearest feedback endpoint under $\rho(x, z) = \min_j d_j(x, z)/\delta_j$ and count a visit when $\rho \leq 1$. Positive and negative groups use the endpoints' saved request-normalized advantages. Table 7 reports geometric visitation rates, with all ordinary outputs in the denominator.

Table 7: **Signed feedback changes local generation.** Geometric visitation rates under shared ordinary sampling noise, using the same first-round feedback endpoints.

Policy	Positive-advantage visits (%)	Negative-advantage visits (%)
Before update	16.11	66.96
ELITE	22.72	52.30
SCOPE	**24.75**	**49.45**

Relative to ELITE, SCOPE increases positive-advantage visitation by 2.04 percentage points and reduces negative-advantage visitation by 2.85 points. Paired 95% Monte Carlo intervals are $[1.88, 2.19]$ and $[-3.02, -2.68]$ points, respectively. Both aggregate directions hold across the two prespecified neighborhood definitions, using the minimum or maximum normalized distance, and all three scales 0.5, 1, and 2. Intervals use 2,000 resamples of shared 128-sample noise blocks within each request, measuring sampling uncertainty for the fitted policies. The same physical feedback thus produces a stronger shift toward higher-advantage neighborhoods when the signed correction is included, complementing the physical coverage results in Table 3.

C.5 Request transfer after controlled exploration

Table 8 extends the SCOPE ablations in Table 3 to eight additional requests. Joint success is similar across variants, at 35.16–37.50%. Relative to Offline FM's seven qualified novel designs, the full configuration and its positive replay ablation (ELITE) each supply 12; uniform replay supplies 11 and current-round priority supplies 17. Replay choices within SCOPE thus affect cross-request reuse as well as generation within the search domain.

Table 8: **Request transfer of SCOPE ablations.** Final ordinary generation on eight additional requests, without further updates. Novel columns count distinct designs beyond $\mathcal{D}_0$ under each criterion.

Policy	Valid (%)	Joint (%)	Novel V	Novel G_τ
Offline FM	**50.39**	37.11	13	7
ELITE	**50.39**	37.11	19	12
SCOPE-Uniform	50.00	35.16	18	11
SCOPE-Current	50.00	37.11	**25**	**17**
SCOPE	50.00	**37.50**	17	12

C.6 STABILITY ACROSS ONLINE RANDOM SEEDS

We repeat the eight-round study in Figure 3 for DFF, ELITE, and SCOPE with a second seed for candidate generation and online training. Both runs use the same pretrained weights, eight requests, physical-evaluation cap, and update budget. Each method starts a separate archive at $\mathcal{D}_0$ and follows the protocol in Appendix A.5. Table 9 reports cumulative discoveries and additions during the final four rounds; V/G_τ counts are subsets of the admitted R representatives.

Table 9: **Online discovery across random seeds.** New representatives added to each method's dynamic archive over all eight rounds and during the final four. Best per seed is bold.

		Rounds 1–8			Rounds 5–8	
Seed	Policy	R	V	G_τ	V	G_τ
62501	DFF	**1098**	845	47	412	20
	ELITE	1074	872	294	428	108
	SCOPE	1095	**954**	**344**	**503**	**148**
62502	DFF	**1099**	838	61	408	34
	ELITE	1026	850	302	416	130
	SCOPE	1078	**935**	**326**	**479**	**142**

SCOPE exceeds ELITE and DFF in request-valid and qualified discoveries in both runs. With the second seed, it exceeds ELITE by 10.0% in valid discoveries and 7.9% in qualified discoveries. The corresponding gains during rounds 5–8 are 15.1% and 9.2%, showing that the advantage persists in the latter half of the search. Both seeds reproduce the ordering of SCOPE, ELITE, and DFF on these two measures as their archives grow.

Final coverage across SCOPE replay ablations. We repeat the two replay ablations of SCOPE on the second online trajectory and evaluate their final policies alongside the full configuration (Table 10). The offline starting weights, eight requests, acquisition rules, and update budgets match Table 3. Final evaluation uses 32 ordinary outputs per request, with the same generation noise across configurations and trajectories. Valid and qualified representatives are independently separated beyond $\mathcal{D}_0$. Changing the online seed tests how sampled feedback and subsequent learning affect the final repertoire.

Table 10: **SCOPE replay ablations across two online trajectories.** Cumulative discoveries and final ordinary coverage on the eight online requests. Best per seed is bold.

		Online discoveries		Final generator		
Seed	Policy	V	G_τ	Joint (%)	Novel V	Novel G_τ
62501	SCOPE-Uniform	890	352	**59.38**	86	43
	SCOPE-Current	**985**	**365**	49.61	108	41
	SCOPE	954	344	50.00	**120**	**51**
62502	SCOPE-Uniform	900	330	**55.08**	92	45
	SCOPE-Current	**952**	**349**	51.95	103	44
	SCOPE	935	326	52.73	**128**	**55**

With the second seed, the full configuration generates 128 valid and 55 qualified novel designs, compared with 103 and 44 under current-round priority and 92 and 45 under uniform replay. Restricting priority to the acquisition round again yields more cumulative discoveries, at 952 valid and 349 qualified designs versus 935 and 326 with sustained priority. Uniform replay retains higher joint success in both trajectories, while sustained priority supplies broader ordinary coverage. These runs show the search–coverage trade-off across replay choices within SCOPE.

D SENSITIVITY OF GEOMETRIC COUNTING

We assess geometric scale sensitivity using the saved ordinary outputs of all six models on the target-domain and request-transfer panels. The generated boundaries and physical qualification labels remain fixed. Starting from the three scales in Appendix B, we apply multipliers $\alpha \in \{0.5, 0.75, 1, 1.25, 1.5, 2\}$. Table 11 scales both novelty relative to $\mathcal{D}_0$ and separation between generated representatives. Table 12 holds the $\mathcal{D}_0$ novelty scales fixed and varies only output separation. Both tables report distinct qualified novel designs within each panel in the standard geometry-identity order.

Table 11: **Joint novelty and separation scale sensitivity.** Distinct qualified novel designs when all three thresholds for novelty relative to $\mathcal{D}_0$ and output separation are multiplied by the column value. Counts use the standard geometry-identity order within each request panel.

Model	0.50	0.75	1.00	1.25	1.50	2.00
Target-domain requests						
DFF	45	45	42	35	29	24
DFF-PCA-Joint	83	71	51	35	25	13
Latent-FM	43	31	23	13	8	6
Raw-FM	54	34	26	19	17	12
Raw-FM + J	102	51	29	18	13	7
SCOPE	**192**	**142**	**102**	**73**	**59**	**30**
Request transfer						
DFF	107	106	97	73	60	34
DFF-PCA-Joint	174	**154**	**109**	**84**	**62**	**38**
Latent-FM	140	95	72	48	35	15
Raw-FM	133	95	59	43	35	20
Raw-FM + J	212	118	68	42	25	6
SCOPE	**222**	133	88	51	36	17

For each setting, we also repeat representative extraction in the standard order, its reverse, and 100 deterministic hash orders. Table 13 gives the resulting count ranges at the original scales. On target-domain requests, SCOPE has the largest qualified repertoire at every tested scale and under all 102 orders in both analyses. At the original scales, its count ranges from 98 to 104, compared with 42 for DFF and 51–52 for DFF-PCA-Joint. The transfer panel retains a distinct scale-dependent pattern, including broader coverage by DFF-PCA-Joint at the original and larger scales.

Table 12: **Output-separation scale sensitivity.** Distinct qualified novel designs with $\mathcal{D}_0$ novelty evaluated at the original scales. The column value multiplies only the thresholds separating generated representatives. Counts use the standard geometry-identity order.

Model	0.50	0.75	1.00	1.25	1.50	2.00
Target-domain requests						
DFF	43	43	42	42	41	37
DFF-PCA-Joint	58	55	51	46	42	31
Latent-FM	27	25	23	23	20	17
Raw-FM	26	26	26	26	26	26
Raw-FM + J	45	33	29	22	19	18
SCOPE	**136**	**116**	**102**	**86**	**79**	**58**
Request transfer						
DFF	102	102	97	89	80	67
DFF-PCA-Joint	**123**	**122**	**109**	**102**	**91**	**74**
Latent-FM	84	79	72	66	62	52
Raw-FM	62	61	59	57	56	54
Raw-FM + J	80	76	68	63	60	48
SCOPE	105	94	88	84	73	66

Table 13: **Representative-counting order sensitivity.** Ranges of distinct qualified novel designs across 102 counting orders at the original geometric scales. A single value indicates the same count under every order.

Model	Target-domain requests	Request transfer
DFF	42	96–97
DFF-PCA-Joint	51–52	108–109
Latent-FM	22–24	69–72
Raw-FM	26	59
Raw-FM + J	28–30	67–70
SCOPE	98–104	85–89

E PHYSICAL ALTERNATIVES AT A SHARED DESIGN REQUEST

Geometric coverage describes the range of equilibria available for subsequent design. Here we connect that range to physical choices within a shared request, using two qualified outputs from the complete SCOPE system.

Physical variation within design families. Stellarator searches explore continuous families of configurations with different magnetic properties. QUASR exhibits such continua within its design clusters (Giuliani et al., 2025). Its analysis of quasi-axisymmetric devices finds visually similar boundary cross-sections; the geometrically close PRECISE_QA and GIU_QA designs differ in QS quality by almost an order of magnitude (Section 5.6 and Figure 14 of that work). Boundary separation and magnetic response therefore provide complementary descriptions of a design repertoire.

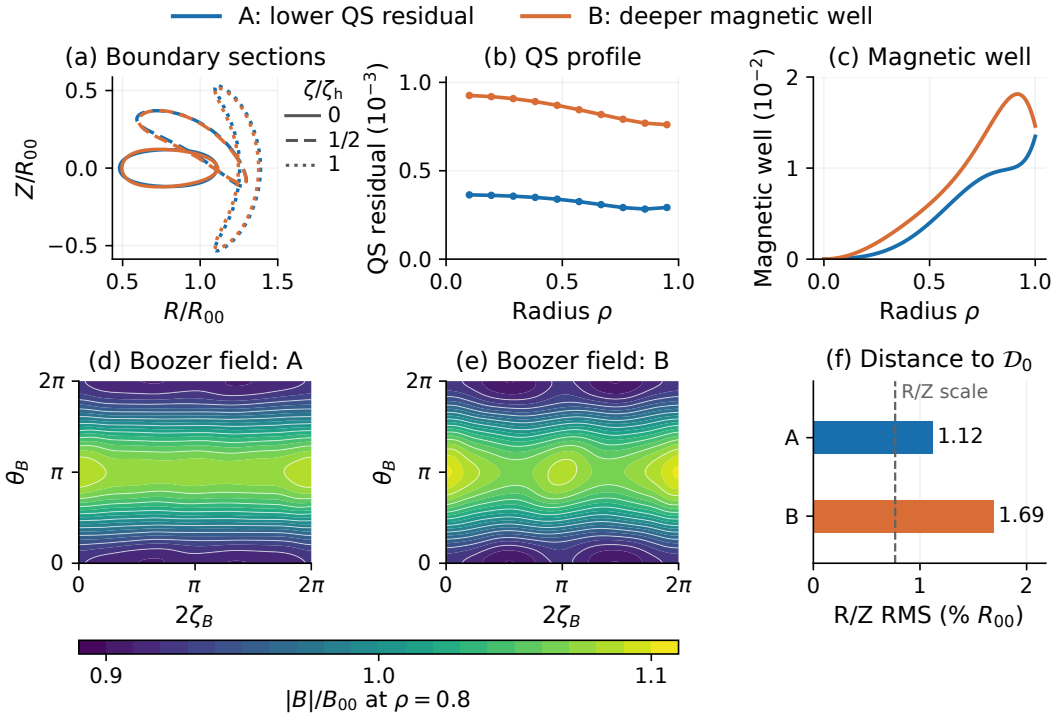


Figure 6: **Physical alternatives under a shared request.** Two qualified SCOPE designs, shown through boundary geometry, radial physical profiles, Boozer field strength, and distance to $\mathcal{D}_0$. Both field maps share their normalization and color scale.

Two qualified realizations of one request. Figure 6 shows ordinary outputs from the target-domain panel in Table 2, at $A^\star = 5.00004$ and $\bar{\iota}^\star = 0.61339$. The two designs share the prescribed pressure, current, magnetic flux, and length scale. Both satisfy $G_\tau = 1$ and belong to the 102 distinct qualified designs beyond $\mathcal{D}_0$. Their boundary R/Z root-mean-square separation is 1.36% of R_{00}; their respective nearest-training-boundary distances are 1.12% and 1.69%. Both also exceed the full-normal and low-frequency novelty scales in Appendix B.

Complementary QS and magnetic-well profiles. Design A has a QS residual of 3.28×10^{-4} and a radial-average magnetic well of 0.00475; Design B has corresponding values of 8.52×10^{-4} and 0.00753. A supplies more precise quasisymmetry, while B supplies a deeper magnetic well throughout the evaluated off-axis profile. The $2.59\times$ QS ratio and $1.59\times$ average-well ratio make their different physical preferences explicit. Their aspect-ratio errors are 0.36% and 0.89%, and their transform errors are 0.00146 and 0.00582, respectively. Thus the additional coverage offers alternative physical realizations while preserving request fidelity and joint qualification.

Case selection and visualization. We select the pair with the smallest boundary separation among same-request qualified novel pairs that are mutually separated and differ by at least $1.5\times$ in both QS residual and average well, with opposing preferences. Boundary sections use equal spatial scales at cylindrical toroidal angles $\zeta/\zeta_{\mathrm{h}} = 0, 1/2, 1$, where $\zeta_{\mathrm{h}} = \pi/N_{\mathrm{FP}} = \pi/2$ is half a field period. Solid, dashed, and dotted lines denote these angles, respectively. Boozer maps are computed from the saved DESC equilibria at $\rho = 0.8$, normalized by each surface's constant field harmonic B_{00}. Their nearly straight contours display the common QA structure; the radial profiles resolve the different quality trade-offs. The normalized force-balance residuals are 4.19×10^{-5} and 4.31×10^{-5} on the working grid, and 3.58×10^{-5} and 3.65×10^{-5} on the denser validation grid. Both designs retain nonnegative well on the 50- and 100-node off-axis radial grids.