SEMANTIC-GUIDED MANIFOLD SEARCH FOR PHYSICAL INVERSE DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Physical inverse design (e.g., spatial coil design) aims to identify geometric configurations satisfying target physical responses. We consider settings with expensive black-box simulations and no access to simulator gradients. As a practically important representative task, spatial coil design optimizes 3D conductive paths for induction heating and magnetic-field generation, and faces challenging high-dimensional geometric search under limited simulation budgets. Vanilla Bayesian optimization does not by itself specify how domain-relevant local deformations should be constructed from simulation feedback. We present Semantic-Guided Manifold Search (SEMGMS), a closed-loop simulation-driven optimization framework. Rather than relying on large-language models to directly generate geometries, SEMGMS distills current simulation feedback and historical trials into structured semantic actions. Its core module, the Geometry Grounding Translator (GGT), converts each semantic action into topology-preserving, bounded local deformation families for graph-represented geometries. A surrogate-guided executor performs Bayesian search strictly within these low-dimensional deformation families. Joint geometric feasibility is checked before each simulation. We validate SEMGMS on SIMCOIL-BENCH, a unified benchmark for spatial-coil inverse design encompassing welding-loss, magnetic-field, and hybrid multi-objective scenarios. Across five-seed evaluations, SEMGMS achieves the lowest mean spatial uniformity errors among the compared methods on all three tasks while maintaining competitive response magnitude.

1 INTRODUCTION

Designing devices for field control and energy delivery requires translating desired physical responses into realizable geometries. Spatial coil design is a representative physical inverse-design problem, where the three-dimensional layout of conductive paths determines the electromagnetic or thermal responses over target regions (Turner, 1986; Kaptanoglu et al., 2024; Tavakoli et al., 2009). It plays an important role in applications such as induction heating, magnetic-field generation and control, and transcranial magnetic stimulation, where coil geometry shapes the magnitude and spatial distribution of the resulting physical responses (Tavakoli et al., 2009; Turner, 1986; Deng et al., 2013; Rissanen et al., 2023). Therefore, coil design is important for delivering sufficiently strong and spatially well-distributed responses while satisfying application-specific target requirements.

Despite its importance, practical spatial coil design remains challenging for three main reasons. *1) Coupled Target Responses*: Geometry changes redistribute responses across target regions, so improving one region may not improve others (Huang & Huang, 2010; Hyun et al., 2025). *2) High-Dimensional Geometric Search*: Localized multipoint responses require flexible spatial deformations, exposing many coupled degrees of freedom to black-box optimization (Kaptanoglu et al., 2024; Rissanen et al., 2023). *3) Limited Simulation Guidance*: Costly physics simulations provide performance observations but no explicit guidance on where, in which direction, or by how much to modify the geometry (Zheng et al., 2025; Shahriari et al., 2016). Consequently, an effective optimizer must infer what geometric operation to apply, where and in which direction to apply it, and how large the modification should be, while preserving geometric feasibility.

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.

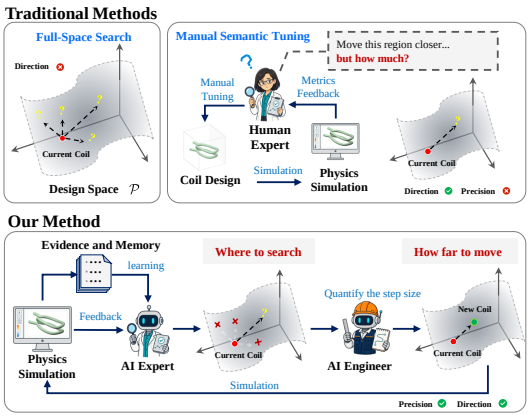


Figure 1: Comparison of full-space search, manual semantic tuning, and SEMGMS. Full-space search lacks directional guidance; manual tuning identifies modifications but cannot determine precise magnitudes. SEMGMS infers search locations and directions from simulation evidence, grounds actions into a feasible deformation space, and determines the step size numerically.

Two existing research directions provide complementary capabilities for addressing these challenges. **1) Numerical Black-Box Optimization.** Derivative-free and Bayesian optimization methods search numerical variables, with high-dimensional variants using random embeddings, additive decompositions, or local trust regions (Snoek et al., 2012; Hansen & Ostermeier, 2001; Wang et al., 2013; Kandasamy et al., 2015; Eriksson et al., 2019). However, optimizing a given parameterization does not by itself specify how localized physical deficiencies should determine coordinated deformation modes. **2) LLM-Guided Optimization.** Recent methods employ LLMs to propose candidate solutions, summarize optimization histories, revise solutions from textual feedback, or assist Bayesian optimization (Yang et al., 2024; Yuksekgonul et al., 2025; Cheng et al., 2024; Liu et al., 2024b; Yuan et al., 2026). LLMs can interpret simulator feedback, but their suggestions do not precisely specify coordinated control-point motions or magnitudes. Direct coordinate generation also requires numerical precision and geometric validity. The challenge is to connect semantic interpretation with numerical execution by turning simulator evidence into an executable geometric search space.

We propose **Sem**antic-**G**uided **M**anifold **S**earch (SEMGMS), a simulation-driven method for spatial coil design. It uses simulator evidence to build action-conditioned geometric reparameterizations instead of asking an LLM to optimize coil geometry directly. A **S**emantic **A**ction **P**lanner (**SAP**) distills current physical responses and trial history into a structured program specifying the operation, focus, direction, and coarse range. A **G**eometry **G**rounding **T**ranslator (**GGT**) maps each action to graph support, coordinated deformation modes, and a bounded coefficient domain. These define an executable, path-aware local deformation space. A **S**urrogate-**G**uided **E**xecutor (**SGE**) searches these spaces to determine precise coefficients from numerical observations. This separation of semantic planning and numerical execution converts simulator evidence into action-bounded spaces, replacing blind full-coordinate exploration with physically meaningful, feasibility-aware search. Our contributions are as follows:

- **Semantic-Guided Manifold Search:** We introduce SEMGMS, a closed-loop framework that reformulates spatial coil optimization as structured semantic action planning followed by geometry-grounded numerical execution. SAP distills simulator responses and accumulated trial history into deformation actions specifying the operation, focus, direction, and coarse range.
- **Geometry-Grounded Numerical Execution:** We develop GGT to map each structured semantic action to a concrete, path-aware graph support, a set of coordinated deformation modes, and a bounded coefficient domain, and SGE to subsequently determine precise deformation coefficients through surrogate-guided search.
- **Unified Benchmark and Comprehensive Evaluation:** We establish SIMCOIL-BENCH, covering Welding Loss Optimization (WLO), Magnetic Field Optimization (MFO), and Hybrid Field-Loss Optimization (HFLO) under a common coil representation and evaluation protocol. Extensive comparisons and ablations show that SEMGMS consistently improves spatial uniformity while maintaining competitive response magnitude.

2 RELATED WORK

Numerical Black-Box Optimization spans derivative-free search and model-based methods that rely only on function evaluations. In AI for science, this setting arises when each query requires a costly experiment or high-fidelity simulation, as in autonomous chemical experimentation, reaction

optimization, and electromagnetic inverse design (Burger et al., 2020; Shields et al., 2021; Zheng et al., 2025). Classical approaches include random search, CMA-ES, direct-search, interpolation, and learned partitions (Bergstra & Bengio, 2012; Hansen & Ostermeier, 2001; Larson et al., 2019; Wang et al., 2020). For expensive simulators, Bayesian optimization allocates limited evaluations through a surrogate and acquisition rule (Snoek et al., 2012); high-dimensional variants use subspaces, trust regions, nested embeddings, or structural priors (Wang et al., 2013; Kandasamy et al., 2015; Eriksson et al., 2019; Jaquier & Rozo, 2020; Papenmeier et al., 2022; Hvarfner et al., 2024; Yu et al., 2026). These methods derive search restrictions from coordinates and objective values rather than translating a localized physical deficiency into an operation, focus, direction, and range. SEMGMS instead uses semantic actions to define topology-preserving, action-bounded deformation families for surrogate search.

LLM-Guided Optimization has been used to optimize prompts and textual objectives through generation, planning, feedback, and trace-based updates (Zhou et al., 2023; Pryzant et al., 2023; Wang et al., 2024; Yang et al., 2024; Chen et al., 2024; Cheng et al., 2024; Yuksekgonul et al., 2025). They serve as mutation, program-generation, or heuristic-design operators in externally evaluated search loops (Chen et al., 2023; Romera-Paredes et al., 2024; Ma et al., 2024; Liu et al., 2024a; Ye et al., 2024). Recent systems couple LLM reasoning with numerical or Bayesian optimization through preferences, acquisition selection, or transfer, including constrained scientific design (Liu et al., 2024b; Chen et al., 2025; Yuan et al., 2026; Ngo et al., 2026; Zeng et al., 2026; Chen et al., 2026; Schwanke et al., 2026). They establish useful search priors but generally operate on text, programs, flat vectors, BO components, or partitions. In contrast, SEMGMS uses the LLM only to infer a semantic deformation action from simulation evidence, while GGT grounds the action into a bounded, path-aware deformation family and SGE determines the deformation coefficients.

3 PROBLEM DEFINITION

We represent a spatial coil by a directed control-point graph $G = (V, E)$, where V contains the control points and E specifies their physical connectivity and routing order. Let $n = |V|$ and $p_i \in \mathbb{R}^3$ denote the position of control point i. The coil geometry is given by $P = [p_1, \dots, p_n]^\top \in \mathbb{R}^{n\times 3}$. During optimization, we keep the graph topology fixed and update only the coordinates of the optimizable control points. Let $V_{\text{fix}} \subseteq V$ index the fixed control points, whose prescribed positions are denoted by $\bar{p}_i$. The feasible design space is $\mathcal{P} = \big\{P \in \mathbb{R}^{n\times 3} \;\big|\; p_i = \bar{p}_i,\ \forall i \in V_{\text{fix}}, g_r(P) \le 0,\ r = 1, \dots, R\big\}$, where g_r denotes the r-th hard geometric or manufacturing constraint. Given a design $P \in \mathcal{P}$, an expensive black-box simulator $\mathcal{S}$ returns $\big(Y(P), C(P)\big) = \mathcal{S}(G, P)$, where $Y(P)$ contains the physical responses over the target regions and $C(P)$ contains constraint measurements handled through soft penalties. No analytical gradients of these responses with respect to P are assumed to be available. We define the scalar objective as

$$F(P) = \mathcal{L}_{\text{task}}\big(Y(P)\big) + \lambda\Phi\big(C(P)\big), \tag{1}$$

where $\mathcal{L}_{\text{task}}$ converts task-specific requirements on response magnitude and spatial uniformity into a scalar loss, Φ penalizes soft constraint violations, and $\lambda \ge 0$ controls the penalty weight. For every benchmark, we use $\mathcal{L}_{\text{task}}(Y(P)) = \max_{1\le m\le M} \ell_m(P)$, where $\ell_m(P)$ is a threshold-referenced signed loss: threshold minus value for maximized metrics, and value minus threshold for minimized metrics, without additional normalization. Given a simulator-evaluation budget B, the optimizer selects a sequence of feasible designs $\{P^{(b)}\}_{b=1}^{B}$ based on previously observed evaluations and returns

$$\widehat{P}_B \in \arg\min_{1\le b\le B} F\big(P^{(b)}\big). \tag{2}$$

The challenge lies in the mismatch between simulation feedback and geometric decision variables. Comparing $Y(P)$ with the task requirements can identify underperforming target regions, but it does not determine which connected control points should move together, what deformation operation should be applied, or what direction and magnitude should be used. Direct full-coordinate optimization must therefore discover these coupled choices through expensive simulator evaluations. This mismatch motivates an action-conditioned reparameterization that specifies the deformation type, affected region, and direction before optimizing a small set of continuous deformation coefficients.

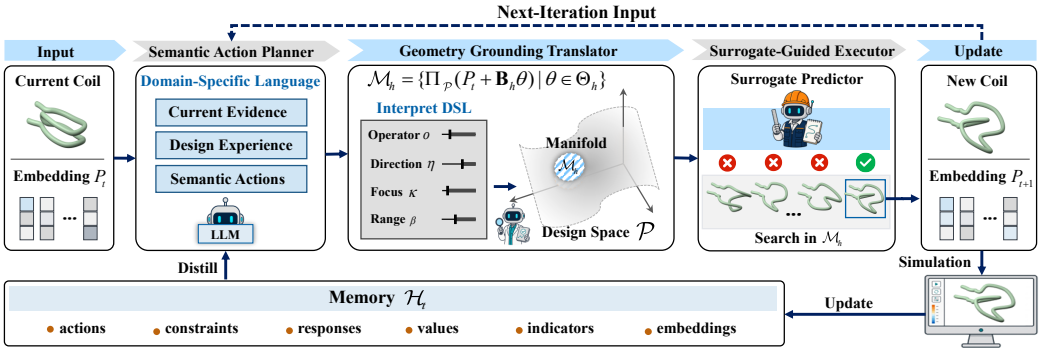


Figure 2: Overview of SEMGMS. At iteration t, SAP distills current evidence and accumulated experience into structured actions specifying the operation, focus, direction, and coarse range. GGT grounds each action on the current coil P_t into a bounded, path-aware deformation family. SGE searches these families using a surrogate fitted to prior evaluations and selects coils for simulation. Simulation outcomes update the evidence memory and determine P_{t+1}.

4 SEMANTIC-GUIDED MANIFOLD SEARCH

SEMGMS separates semantic planning from geometry-grounded numerical execution in a simulation-driven closed loop, as illustrated in Figure 2 and summarized in Algorithm 1. SAP distills current responses and trial history into constrained actions specifying operation, focus, direction, and coarse range. GGT grounds each action on the current coil as a bounded, path-aware family with graph support and coordinated modes. SGE fits a surrogate to prior evaluations, searches these families, and updates evidence memory with simulation outcomes.

Algorithm 1 Pipeline of the proposed SEMGMS

Require: Graph $G = (V, E)$, initial embedding P_{init}, simulator $\mathcal{S}$, objective F, operator library $\mathcal{O}$, budgets B_0, T, K with $B = B_0 + TK$
Ensure: Best feasible coil embedding P_{best}

```
Initialize H_0 with B_0 evaluations seeded by P_init
P_best ← arg min_{P_i ∈ H_0} F(P_i); P_0 ← P_best
for t = 0 to T − 1 do
    Read cached (Y_t, C_t) at P_t from H_t
    E_t ← φ_mem(H_t); s_t ← (E_t, Y_t, C_t, G)
    δ_t ← π_SAP(s_t; O)
    D_t ← τ_DSL(δ_t; O)
    M_t ← ∅
    for h ∈ D_t do
        (S_h, B_h, Θ_h) ← τ_GGT(h; G, P_t)
        M_h ← {Π_P(P_t + B_h θ) | θ ∈ Θ_h}
        M_t ← M_t ∪ M_h
    end for
    Fit surrogate μ_t to I_t
    Select {(h_j, θ_j)}_{j=1}^K via α_t(ψ_h(θ); μ_t)
    Form Q_t = {(ψ_{h_j}(θ_j), h_j)}_{j=1}^K
    Add Q_t evaluations to H_t, yielding H_{t+1}
    Update P_best; choose anchor P_{t+1}
end for
return P_best
```

4.1 SEMANTIC ACTION PLANNER

Simulator evaluations expose physical deficiencies but do not identify the geometric action needed to correct them. To obtain a reusable search direction, SAP distills current responses and prior trials into a constrained semantic-action program. Let $\mathcal{H}_t$ denote the design evidence memory available at iteration t: $\mathcal{H}_t = \{(P_i, h_i, Y_i, C_i, F_i, z_i)\}_{i=1}^{N_t}$. Here, N_t is the number of simulator-evaluated designs in memory. The variable h_i is the semantic action from which P_i was sampled when applicable. Specifically, $Y_i = Y(P_i)$, $C_i = C(P_i)$, $F_i = F(P_i)$, and z_i records whether P_i improved the incumbent. Let φ_{mem} denote the LLM-based evidence-distillation map, which compresses this history into reusable design experience $\mathcal{E}_t = \varphi_{\text{mem}}(\mathcal{H}_t)$. Together with the current anchor responses $Y_t = Y(P_t)$ and $C_t = C(P_t)$, this summary forms the evidence state $\mathbf{s}_t = (\mathcal{E}_t, Y_t, C_t, G)$. The state contains target-region residuals, feasibility signals, distilled action outcomes, and, when available, graph-local responsibility information. Conditioned on $\mathbf{s}_t$, SAP uses the LLM to emit a constrained semantic-action DSL program

$$\delta_t = \pi_{\text{SAP}}(\mathbf{s}_t; \mathcal{O}), \tag{3}$$

where π_{SAP} denotes the semantic planning policy and $\mathcal{O}$ is the deformation-operator library. Each DSL clause encodes $h = (o, \kappa, \eta, \beta)$: an operator, focus, qualitative direction, and coarse range. SAP uses evidence to decide what kind of coil modification is physically plausible, but it does not

generate new coordinates, identify active coordinate entries, or choose displacement coefficients. For example, when the loss around tube B is too low, SAP may request a local translation of a responsible coil region toward B within a conservative range. When a region is close to a clearance or curvature violation, SAP may instead request smoothing or relaxation. Thus, SAP specifies the operation, targeted physical issue, coarse focus, direction, and range, leaving exact graph support, geometric parameterization, and numerical displacement to subsequent stages.

4.2 Geometry Grounding Translator

A semantic action specifies what and where to modify, but remains a language-level instruction that cannot be applied to the coil embedding. GGT resolves this ambiguity by mapping each DSL clause to path-aware deformation modes and bounded coefficients on the current geometry. Given the DSL program δ_t, GGT first validates it against the typed operator library and converts it into an ordered action list $\mathcal{D}_t = \tau_{\mathrm{DSL}}(\delta_t; \mathcal{O}) = (h_{t,1}, \ldots, h_{t,L_t})$. Here, τ_{DSL} is the typed DSL translation map and L_t is the number of valid clauses. Each valid action $h \in \mathcal{D}_t$ is then mapped by $\tau_{\mathrm{GGT}}(h; G, P_t)$ to three objects: a graph support S_h, a deformation basis $\mathbf{B}_h$, and a bounded coefficient domain Θ_h.

GGT first resolves the focus κ on the current embedding P_t. This produces a concrete graph support $S_h \subseteq V$, i.e., the subset of control points whose coupled motion realizes the semantic intent. The support may extend around κ to maintain smooth deformation and directed connectivity.

Given the support S_h, GGT constructs d_h local deformation modes $d_h^{(j)} = [d_{h,1}^{(j)}, \ldots, d_{h,n}^{(j)}]^\top \in \mathbf{R}^{n\times 3}$. Each mode follows an operator-specific displacement field inside S_h, with $d_{h,v}^{(j)} = w_{h,v} a_{h,v}^{(j)}(P_t, o, \eta)$ for $v \in S_h$, and is zero otherwise, where $w_{h,v} \in [0,1]$ is a locality weight and $a_{h,v}^{(j)}$ is the displacement direction or update field induced by the operator o and direction η. They form the deformation basis,

$$\mathbf{B}_h = (d_h^{(1)}, \ldots, d_h^{(d_h)}), \qquad \mathbf{B}_h\theta = \sum_{j=1}^{d_h} \theta_j d_h^{(j)}. \tag{4}$$

The basis $\mathbf{B}_h$ defines coupled graph-level motions rather than independent coordinate perturbations. For example, a directional translation action induces displacement modes over S_h, while smoothing, layer-offset, or redistribution actions induce modes that move neighboring control points jointly.

GGT then converts the semantic range β and local feasibility margins into the bounded coefficient domain Θ_h. The semantic range gives per-mode bounds $b_{h,j}(\beta)$, while feasibility induces the largest locally admissible step $m_{h,j} = \sup\{r \geq 0 : P_t + s d_h^{(j)} \in \mathcal{P} \text{ for all } |s| \leq r\}$. With safety factor $\gamma \in (0,1]$, the executable bound is $\bar{\theta}_{h,j} = \min\{b_{h,j}(\beta), \gamma m_{h,j}\}$. The bounded action-variable domain is $\Theta_h = \{\theta \in \mathbf{R}^{d_h} : |\theta_j| \leq \bar{\theta}_{h,j},\ j = 1, \ldots, d_h\}$.

Grounding defines the map and deformation family:

$$\psi_h(\theta) = \Pi_{\mathcal{P}}(P_t + \mathbf{B}_h\theta), \quad \mathcal{M}_h = \{\psi_h(\theta) : \theta \in \Theta_h\}. \tag{5}$$

When $\Pi_{\mathcal{P}}$ is inactive, the interior of $\mathcal{M}_h$ is a manifold patch. Otherwise, we use the general term deformation family. The per-mode bounds conservatively limit each deformation direction but do not by themselves guarantee that every multimode combination is feasible. Accordingly, $\Pi_{\mathcal{P}}$ performs the joint feasibility check, either projecting a design back to the feasible set or rejecting it when hard path-feasibility constraints are violated. Thus, GGT does not perform simulator evaluation and does not choose the final action coefficients. Its role is to turn a typed semantic action into a bounded, path-aware deformation family for SGE.

4.3 Surrogate-Guided Executor

Grounding defines admissible modes and bounded coefficients; their physical effects remain accessible only through expensive black-box simulation. SGE therefore fits a surrogate to evaluated designs and searches the grounded coefficient domains:

$$\mathcal{M}_t = \bigcup_{h\in\mathcal{D}_t} \mathcal{M}_h, \quad \min_{P\in\mathcal{M}_t} F(P) = \min_{\substack{h\in\mathcal{D}_t \\ \theta\in\Theta_h}} F(\psi_h(\theta)). \tag{6}$$

It extracts the embedding-objective observations from $\mathcal{H}_t$:

$$\mathcal{I}_t = \{(P_i, F_i)\}_{i=1}^{N_t}, \tag{7}$$

where $F_i = F(P_i)$. Let μ_t denote a surrogate predictor fitted on the observations in Equation (7), and let $\alpha_t(P; \mu_t)$ be the acquisition score. SGE ranks action-coefficient pairs by $\alpha_t(\psi_h(\theta); \mu_t)$ over $h \in \mathcal{D}_t$ and $\theta \in \Theta_h$, takes the top K, and maps them back to coil embeddings: $\mathcal{Q}_t = \{(\psi_{h_j}(\theta_j), h_j)\}_{j=1}^{K}$. Here, K is the number of simulator queries issued in the current execution step. The search set $\mathcal{M}_t$ defines where the acquisition rule is allowed to search, whereas $\mathcal{I}_t$ provides the numerical supervision used to learn the surrogate. The evidence memory $\mathcal{H}_t$ remains available when summarizing experience and planning the next action. Instead of optimizing over the full embedding space, SGE queries the surrogate only on designs instantiated by bounded action coefficients within the current semantic actions.

Each query-action pair $(P', h') \in \mathcal{Q}_t$ uses a chosen action and bounded coefficients to obtain $P' = \psi_h(\theta)$. Only these designs are evaluated by $\mathcal{S}$; responses, constraint measurements, objectives, and improvement indicators are appended to $\mathcal{H}_{t+1}$. Successful actions provide positive planning evidence, while failed or infeasible actions provide negative evidence. Algorithm 1 summarizes the loop: the LLM selects the semantic region, GGT makes it executable, and SGE searches it. SEMGMS searches each grounded deformation family using only evaluated design–objective pairs, without simulator derivatives. For local analysis, we project the gradient of a smooth branch of F at P_t onto $\mathrm{Range}(B_h)$, quantifying the descent information captured by the grounded span separately from SGE's finite-budget search error.

4.4 THEORETICAL MOTIVATION

For this analysis, identify the D free coordinates of P_t with a vector in $\mathbb{R}^D$. Let $\mathbf{B}_h \in \mathbb{R}^{D \times d_h}$ collect the corresponding deformation modes. Let $\bar{d}_h = \mathrm{rank}(\mathbf{B}_h)$ and let $\mathbf{Q}_h$ be an orthonormal basis of $\mathrm{Range}(\mathbf{B}_h)$. At anchors where the active maximum and penalty branches are locally unchanged, write $\mathbf{g}_t = \nabla F(P_t) \neq 0$ and define the alignment score

$$\rho_h = \frac{\|\mathbf{Q}_h^\top \mathbf{g}_t\|_2^2}{\|\mathbf{g}_t\|_2^2} \in [0, 1]. \tag{8}$$

For SGE's best feasible query $\widehat{P}_t$, let $\varepsilon_t = F(\widehat{P}_t) - \inf_{P \in \mathcal{M}_t} F(P) \geq 0$.

Proposition 1 (Conditional local descent)**.** *Assume that F is L-smooth near P_t, $\Pi_{\mathcal{P}}$ is locally inactive, and $\mathcal{M}_h$ contains a radius-ν_h relative ball around P_t in* $\mathrm{Range}(\mathbf{B}_h)$*. If $\rho_h > 0$ and $0 < \alpha \leq \min\{1/L, \nu_h / \|\mathbf{Q}_h \mathbf{Q}_h^\top \mathbf{g}_t\|_2\}$, then*

$$F(\widehat{P}_t) \leq F(P_t) - \alpha \left(1 - \frac{L\alpha}{2}\right) \rho_h \|\mathbf{g}_t\|_2^2 + \varepsilon_t. \tag{9}$$

Lemma 1 (Random-subspace reference)**.** *For fixed $\mathbf{g}_t$, an independently sampled Haar-uniform rank-$\bar{d}_h$ subspace satisfies $\mathbb{E}[\rho_{\mathrm{rand}}] = \bar{d}_h / D$.*

Proposition 1 separates two factors: ρ_h measures how well the grounded modes align with local descent, while ε_t measures SGE's finite-budget search error. Lemma 1 gives the same-rank random reference $\bar{d}_h / D$. An evidence-guided family with $\rho_h > \bar{d}_h / D$ retains more local descent signal than this reference. The bound yields a decrease whenever its descent term exceeds ε_t. This provides local motivation, not a global convergence guarantee. See the appendix for proofs.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETTINGS

Benchmarks. To comprehensively evaluate spatial coil optimization methods, we introduce SIMCOIL-BENCH, a simulation-driven benchmark suite comprising three complementary tasks across diverse physical objectives and target geometries. To the best of our knowledge, it is the first unified benchmark to evaluate localized welding-loss, volumetric magnetic-field, and coupled field and loss optimization under a common graph-based coil representation and simulator-evaluation

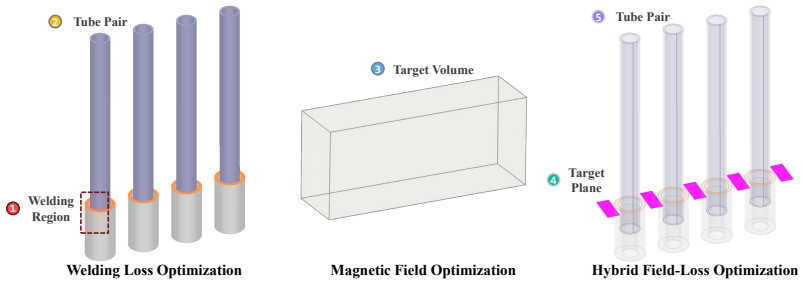


Figure 3: Overview of the three SIMCOIL-BENCH tasks. WLO evaluates welding-plane induced loss and inter-pair loss uniformity, MFO evaluates magnetic-field strength and uniformity over a target volume, and HFLO jointly evaluates welding-loss and magnetic-field responses.

Table 1: Performance over five random seeds on the three SIMCOIL-BENCH tasks, with 200 simulator evaluations per run. Each entry reports the mean ± standard deviation; uniformity errors are in percent. Bold denotes the best mean.

	WLO		MFO		HFLO		
Method	Loss ↑	Loss Unif. ↓	Field Str. ↑	Field Unif. ↓	Loss ↑	Loss Unif. ↓	Field Unif. ↓
Initial Design	110.69±5.15	17.40±2.11	100.34±2.11	62.62±7.10	97.28±1.23	25.84±0.83	25.92±0.33
Random	112.77±3.90	9.08±0.14	96.44±2.74	31.74±4.06	124.78±1.61	15.89±0.67	13.07±0.59
TuRBO	118.59±0.93	5.36±0.23	99.28±1.35	24.71±1.15	170.84±1.56	7.35±0.52	6.95±0.11
SAASBO	121.50±2.20	7.09±0.23	**105.90**±2.31	25.60±2.55	165.63±0.92	7.53±0.12	6.70±0.74
OPRO	165.73±6.01	6.99±0.23	98.95±2.11	25.94±1.76	136.78±5.64	10.97±1.43	10.33±1.44
TextGrad	146.20±6.63	6.09±0.81	102.94±1.98	27.49±2.70	136.02±5.80	10.61±0.74	10.23±0.76
LLAMBO	137.73±9.25	4.91±0.46	101.21±1.41	29.09±1.95	146.87±2.77	12.10±1.78	11.28±1.49
HOLLM	138.36±3.12	7.46±0.53	102.59±1.62	27.93±2.14	134.90±3.47	11.08±1.05	9.36±0.63
SEMGMS	**166.36**±2.52	**3.96**±0.29	104.32±2.23	**23.45**±1.07	**171.00**±1.44	**5.95**±0.49	**5.57**±0.42

protocol. **1) Welding Loss Optimization (WLO)** evaluates localized energy delivery by maximizing the induced-loss strength at each welding plane while reducing imbalance across tube pairs. **2) Magnetic Field Optimization (MFO)** raises the task from discrete welding regions to a spatial target volume, jointly evaluating magnetic-field strength and uniformity throughout the volume. **3) Hybrid Field-Loss Optimization (HFLO)** couples the welding-loss and magnetic-field objectives in a multi-objective setting. It tests whether an optimizer can allocate its search focus across heterogeneous and potentially competing physical feedback, rather than improving one response at the expense of the other. For all tasks, each run starts from a randomly generated coil embedding with fixed connectivity. The inlet and outlet power-interface paths remain fixed, while the remaining coil body is optimized. All experiments use the default two-layer configuration and a budget of 200 simulator evaluations. Figure 3 illustrates the target geometries of the three tasks.

Metrics. We evaluate each benchmark in terms of response magnitude and spatial uniformity. For WLO, we report induced loss (**Loss** ↑) and range-based loss-uniformity error (**Loss Unif.** ↓), following commonly used power-density, Ohmic-loss, and heating-uniformity criteria in induction-heating design (Spateri et al., 2024; Choubey et al., 2025; Uzzell et al., 2024). For MFO, we report magnetic-field strength within the target volume (**Field Str.** ↑) and corresponding range-based field-uniformity error (**Field Unif.** ↓) (Turner, 1986; Wu et al., 2019). For HFLO, we report induced loss (**Loss** ↑), loss-uniformity error (**Loss Unif.** ↓), and field-uniformity error (**Field Unif.** ↓), measuring welding-region energy delivery, inter-pair balance, and target-volume field uniformity, respectively.

Baselines. We compare SEMGMS with three categories of optimization baselines and include the randomly generated **Initial Design** as a no-optimization reference. **1) Classical optimizers** include Random Search, TuRBO (Eriksson et al., 2019), and SAASBO (Eriksson & Jankowiak, 2021), covering unguided sampling and high-dimensional Bayesian optimization. **2) LLM-as-optimizer methods** include OPRO (Yang et al., 2024) and TextGrad (Yuksekgonul et al., 2025), which are adapted to propose coil modifications directly from simulation feedback. **3) LLM-enhanced opti-**

Table 2: Ablation of GGT on WLO. SAP emits the complete semantic-action DSL in all variants, while GGT replaces the action-conditioned focus and/or range with fixed, action-independent defaults.

GGT Variant	Loss ↑	Unif. Err. ↓
Fixed Focus and Range	**170.00**	5.39
Fixed Action Range	160.98	4.25
Fixed Action Focus	157.81	4.05
SEMGMS	167.07	**3.62**

Table 3: Ablation of SGE on WLO. We replace surrogate-guided query ranking with random sampling or LLM-based execution and separately remove surrogate initialization, while keeping SAP and GGT unchanged.

SGE Variant	Loss ↑	Unif. Err. ↓
Random Sampling	147.69	9.03
w/o Surrogate Initialization	156.09	6.13
LLM-Based Execution	120.75	7.14
SEMGMS	**167.07**	**3.62**

mizers include LLAMBO (Liu et al., 2024b) and HOLLM (Schwanke et al., 2026), which integrate LLM reasoning with numerical or Bayesian optimization.

Implementation Details. For each benchmark, we evaluate five random seeds with different initial coils. All methods share the initial coil within each seed and receive 200 simulator evaluations per run. Each repeated run independently samples the stochastic LLM decoding process in addition to numerical optimization randomness. SEMGMS uses the first 100 evaluations to initialize its surrogate and evidence memory, and the remaining 100 evaluations for semantic-guided optimization. It uses DeepSeek-V4-Flash with temperature 0.7, top-p 1.0, and a maximum output length of 8000 tokens. The numerical executor models the scalar utility $U(P) = -F(P)$ with a BoTorch SingleTaskGP and maximizes qLogExpected Improvement in U, evaluating one design per iteration. All LLM-based methods use the same backbone, while their decoding settings follow the respective optimization procedures. All ablation experiments use random seed 0.

5.2 COMPARISON EXPERIMENTS

As shown in Table 1, across five random seeds with 200 simulator evaluations per run, SEMGMS achieves the lowest mean uniformity errors on all SIMCOIL-BENCH tasks and the highest mean welding loss on WLO and HFLO.

SEMGMS achieves the best mean loss and uniformity on WLO. SEMGMS obtains the highest mean welding loss of 166.36 ± 2.52 and the lowest mean loss-uniformity error of $3.96 \pm 0.29\%$, improving over the strongest baseline by 0.38% and 19.35%, respectively. The simultaneous gains are consistent with SAP identifying focused deformation actions from regional response deficiencies, while GGT and SGE convert these actions into coordinated and precisely scaled geometric updates.

SEMGMS achieves the best mean field uniformity with competitive field strength on MFO. SEMGMS achieves a mean field-uniformity error of $23.45 \pm 1.07\%$, 5.10% lower than the best baseline mean, while maintaining a mean field strength of 104.32 ± 2.23.

SEMGMS achieves the best means across all HFLO metrics. SEMGMS obtains the highest mean welding loss of 171.00 ± 1.44 and the lowest mean loss- and field-uniformity errors of $5.95 \pm 0.49\%$ and $5.57 \pm 0.42\%$, respectively. Compared with the strongest baseline means, the two uniformity errors are reduced by 19.08% and 16.84%, respectively. These results suggest that SAP can identify useful actions from heterogeneous physical feedback, while GGT and SGE constrain and refine the geometric updates to jointly improve field and lsoss responses.

5.3 ABLATION STUDIES

We ablate three modules of SEMGMS on WLO: GGT action-conditioned grounding, SGE numerical execution, and SAP semantic planning. All variants use the main experimental setup.

GGT grounding. As shown in Table 2, replacing either the action-conditioned focus or range with a fixed default increases uniformity error and reduces welding loss. Fixing both increases welding loss but substantially worsens uniformity, showing that GGT is important for localizing and bounding each action rather than merely instantiating its operator.

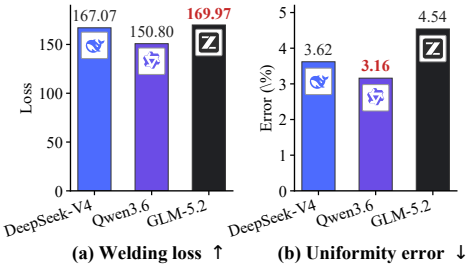


Figure 4: Effect of the LLM backbone on WLO. DeepSeek-V4-Flash, Qwen3.6-Plus, and GLM-5.2 are evaluated with all other components and the 200-evaluation budget fixed.

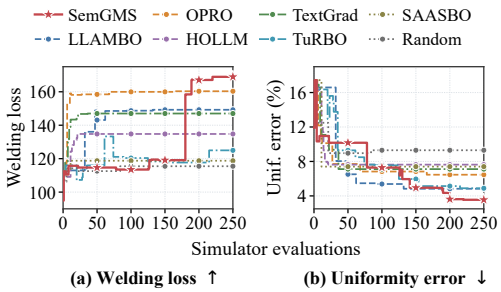


Figure 5: Optimization trajectories on WLO Case 1 with a 250-evaluation budget. Both panels show the same incumbent selected by each method's scalar objective.

SGE execution. All alternatives in Table 3 underperform SEMGMS. LLM-based execution yields the lowest welding loss, while random sampling and removing surrogate initialization substantially increase uniformity error. These results distinguish the modules: SAP identifies meaningful actions, GGT grounds them into localized, bounded deformation families, and SGE resolves their magnitude.

Table 4: Ablation of SAP on WLO. Random Action replaces evidence-conditioned planning with random actions. Other variants alter SAP evidence.

SAP Variant	Loss ↑	Unif. Err. ↓
Random Action	111.60	10.11
w/o Current Feedback	129.54	9.44
w/o Domain Prior	162.02	3.99
Current Feedback Only	158.19	7.82
Distilled Memory Only	155.71	4.88
SEMGMS	**167.07**	**3.62**

SAP planning. As shown in Table 4, Random Action produces the weakest result, confirming that valid syntax alone is insufficient without evidence-conditioned planning. Removing current simulator feedback also sharply degrades both metrics, while using only current feedback or distilled memory underperforms their combination. Domain priors provide a smaller but measurable benefit.

5.4 FURTHER ANALYSIS

LLM backbone sensitivity. To assess whether SEMGMS depends on a specific LLM, we replace its LLM while keeping all other settings fixed. From Figure 4, GLM-5.2 attains the highest loss of 169.97, Qwen3.6-Plus achieves the lowest uniformity error of 3.16%, and DeepSeek-V4-Flash provides a balanced result of 167.07 and 3.62%. The results indicate that SEMGMS is effective across LLMs, although LLM selection affects the balance between welding loss and uniformity.

Optimization dynamics. To examine how performance evolves under the evaluation budget, we extend WLO Case 1 from 200 to 250 evaluations and track the incumbent welding loss and uniformity error. In Figure 5, SEMGMS continues to improve both metrics during later evaluations, whereas most baselines plateau earlier. This sustained progress suggests that the action-conditioned search has not saturated within 250 evaluations and may achieve gains with a larger budget.

6 CONCLUSION

In this paper, we propose **Sem**antic-**G**uided **M**anifold **S**earch (SEMGMS), a simulation-driven framework that converts simulator evidence into action-conditioned local reparameterizations of the geometric search space. Rather than directly generating coil coordinates with an LLM, SAP extracts structured deformation actions from current physical responses and accumulated trial history, GGT grounds them into bounded, path-aware deformation families, and SGE determines precise deformation coefficients through surrogate-guided search. Experiments on SIMCOIL-BENCH, covering WLO, MFO, and HFLO, show that SEMGMS achieves the lowest spatial uniformity error across all three tasks while maintaining competitive or superior response magnitude.

AI USE STATEMENT

Generative AI assistance in preparing this manuscript was limited to language polishing, including improvements to grammar, readability, and phrasing. It was not used to formulate the research problem, develop the proposed methodology, derive theoretical results, implement the research methods, design experiments, or interpret the results. The LLM components evaluated as part of SEMGMS are described in the method and experimental sections. The authors take full responsibility for the accuracy and integrity of the manuscript, including all AI-assisted edits.

ETHICS STATEMENT

This work studies simulation-driven optimization of spatial coil geometries. It does not involve human participants, personal data, or animal experiments. We do not identify specific ethical concerns arising from the research reported here. The reported results are based on numerical simulations; deployment in physical systems would require separate engineering validation and appropriate safety assessment.

REPRODUCIBILITY STATEMENT

To support reproducibility, we provide the theoretical assumptions and derivations in Appendix A, detailed module descriptions and the semantic-action DSL in Appendix B, and simulator settings, benchmark geometries, evaluation metrics, and feasibility constraints in Appendix C. Appendix D specifies the computing environment, random seed settings, model and optimization configurations, evaluation budgets, baseline settings, and LLM decoding protocol. Complete LLM prompt templates and representative outputs are provided in Appendix G. These materials document the experimental setup and procedures used to obtain the reported results. The code for our method and experiments will be open-sourced upon acceptance of the paper.

REFERENCES

Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 20577–20612, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/419f72cbd568ad62183f8132a3605a2a-Abstract-Conference.html.

Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. BoTorch: A framework for efficient monte-carlo bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(10):281–305, 2012. URL https://jmlr.org/papers/v13/bergstra12a.html.

Benjamin Burger, Phillip M. Maffettone, Vladimir V. Gusev, Catherine M. Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M. Alston, Buyi Li, Rob Clowes, Nicola Rankin, Brandon Harris, Reiner Sebastian Sprick, and Andrew I. Cooper. A mobile robotic chemist. *Nature*, 583(7815): 237–241, 2020. doi: 10.1038/s41586-020-2442-2. URL https://doi.org/10.1038/s41586-020-2442-2.

Angelica Chen, David Dohan, and David So. EvoPrompting: Language models for code-level neural architecture search. In *Advances in Neural Information Processing Systems*, volume 36, pp. 7787–7817. Curran Associates, Inc., 2023. doi: 10.52202/075280-0342. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/184c1e18d00d7752805324da48ad25be-Abstract-Conference.html.

Angelica Chen, Samuel Don Stanton, Frances Ding, Robert G Alberstein, Andrew Martin Watkins, Richard Bonneau, Vladimir Gligorijevic, Kyunghyun Cho, and Nathan C. Frey. Generalists vs. specialists: Evaluating LLMs on highly-constrained biophysical sequence optimization tasks. In *Proceedings of the International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 9029–9072. PMLR, 2025. URL https://proceedings.mlr.press/v267/chen25bg.html.

Lichang Chen, Jiuhai Chen, Tom Goldstein, Heng Huang, and Tianyi Zhou. InstructZero: Efficient instruction optimization for black-box large language models. In *Proceedings of the International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 6503–6518. PMLR, 2024. URL https://proceedings.mlr.press/v235/chen24e.html.

Zhuo Chen, Xinzhe Yuan, Jianshu Zhang, Jinzong Dong, Ruichen Zhou, Yingchun Niu, Tianhang Zhou, Yu Yang Fredrik Liu, Yuqiang Li, Nanyang Ye, and Qinying Gu. LABO: LLM-accelerated bayesian optimization through broad exploration and selective experimentation. In *Proceedings of the International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/64679.

Ching-An Cheng, Allen Nie, and Adith Swaminathan. Trace is the next autodiff: Generative optimization with rich feedback, execution traces, and llms. In *Advances in Neural Information Processing Systems*, volume 37, pp. 71596–71642. Curran Associates, Inc., 2024. doi: 10.52202/079017-2287. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/83ba7056bce2c3c3c27e17397cf3e1f0-Paper-Conference.pdf.

Rahul Kumar Choubey, Mayur Patil, and Prashant K. Jain. Tailoring coil geometry for achieving uniform heating through coil shape topology optimization for induction heating-based metal wire additive manufacturing. *Rapid Prototyping Journal*, 31(2):327–343, January 2025. doi: 10.1108/RPJ-08-2024-0360. URL https://doi.org/10.1108/RPJ-08-2024-0360.

Norman Cliff. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3):494–509, 1993. doi: 10.1037/0033-2909.114.3.494. URL https://doi.org/10.1037/0033-2909.114.3.494.

Zhi-De Deng, Sarah H. Lisanby, and Angel V. Peterchev. Electric field depth–focality tradeoff in transcranial magnetic stimulation: Simulation comparison of 50 coil designs. *Brain Stimulation*, 6(1):1–13, 2013. doi: 10.1016/j.brs.2012.02.005.

David Eriksson and Martin Jankowiak. High-dimensional bayesian optimization with sparse axis-aligned subspaces. In Cassio de Campos and Marloes H. Maathuis (eds.), *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pp. 493–503. PMLR, 2021. URL https://proceedings.mlr.press/v161/eriksson21a.html.

David Eriksson, Michael Pearce, Jacob Gardner, Ryan D. Turner, and Matthias Poloczek. Scalable global optimization via local bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 32, pp. 5497–5508. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/6c990b7aca7bc7058f5e98ea909e924b-Abstract.html.

Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, June 2001. doi: 10.1162/106365601750190398. URL https://doi.org/10.1162/106365601750190398.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. doi: 10.2307/4615733. URL https://www.jstor.org/stable/4615733.

Ming-Shyan Huang and Yao-Lin Huang. Effect of multi-layered induction coils on efficiency and uniformity of surface heating. *International Journal of Heat and Mass Transfer*, 53(11–12): 2414–2423, 2010. doi: 10.1016/j.ijheatmasstransfer.2010.01.042. URL https://doi.org/10.1016/j.ijheatmasstransfer.2010.01.042.

Carl Hvarfner, Erik Orm Hellsten, and Luigi Nardi. Vanilla Bayesian optimization performs great in high dimensions. In *Proceedings of the International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 20793–20817. PMLR, 2024. URL https://proceedings.mlr.press/v235/hvarfner24a.html.

Jaeyub Hyun, Darun Barazanchy, Jaspreet Pandher, Michel van Tooren, and H. Alicia Kim. Gradient-based shape optimization of induction welding coil for thermoplastic composites via level-set method. *Composites Part B: Engineering*, 306:112824, 2025. doi: 10.1016/j.compositesb.2025.112824. URL https://doi.org/10.1016/j.compositesb.2025.112824.

Noémie Jaquier and Leonel Rozo. High-dimensional Bayesian optimization via nested Riemannian manifolds. In *Advances in Neural Information Processing Systems*, volume 33, pp. 20939–20951. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f05da679342107f92111ad9d65959cd3-Abstract.html.

Kirthevasan Kandasamy, Jeff Schneider, and Barnabas Poczos. High dimensional bayesian optimisation and bandits via additive models. In Francis Bach and David Blei (eds.), *Proceedings of the International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 295–304, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/kandasamy15.html.

Alan A. Kaptanoglu, Gabriel P. Langlois, and Matt Landreman. Topology optimization for inverse magnetostatics as sparse regression: Application to electromagnetic coils for stellarators. *Computer Methods in Applied Mechanics and Engineering*, 418:116504, 2024. doi: 10.1016/j.cma.2023.116504.

William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952. doi: 10.1080/01621459.1952.10483441. URL https://doi.org/10.1080/01621459.1952.10483441.

Jeffrey Larson, Matt Menickelly, and Stefan M. Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019. doi: 10.1017/S0962492919000060. URL https://doi.org/10.1017/S0962492919000060.

Fei Liu, Xialiang Tong, Mingxuan Yuan, Xi Lin, Fu Luo, Zhenkun Wang, Zhichao Lu, and Qingfu Zhang. Evolution of heuristics: Towards efficient automatic algorithm design using large language model. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32201–32223. PMLR, 2024a. URL https://proceedings.mlr.press/v235/liu24bs.html.

Tennison Liu, Nicolás Astorga, Nabeel Seedat, and Mihaela van der Schaar. Large language models to enhance bayesian optimization. In *International Conference on Learning Representations*, 2024b. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/84b8d9fcb4e262fcd429544697e1e720-Abstract-Conference.html.

Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Jim Fan, and Anima Anandkumar. Eureka: Human-level reward design via coding large language models. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/70c26937fbf3d4600b69a129031b66ec-Abstract-Conference.html.

H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947. doi: 10.1214/aoms/1177730491. URL https://doi.org/10.1214/aoms/1177730491.

Giang Ngo, Dat Phan Trong, Dang Nguyen, Sunil Gupta, and Svetha Venkatesh. Adaptive acquisition selection for bayesian optimization with large language models. In *International Conference on Learning Representations*, 2026. URL https://iclr.cc/virtual/2026/poster/10010684.

Leonard Papenmeier, Luigi Nardi, and Matthias Poloczek. Increasing the scope as you learn: Adaptive Bayesian optimization in nested subspaces. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11586–11601. Curran Associates, Inc., 2022. doi: 10.52202/068431-0842. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/4b7439a4ab0b8e4bcb4e2412c6a10a58-Abstract-Conference.html.

Reid Pryzant, Dan Iter, Jerry Li, Yin Tat Lee, Chenguang Zhu, and Michael Zeng. Automatic prompt optimization with "gradient descent" and beam search. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 7957–7968. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.emnlp-main.494. URL https://aclanthology.org/2023.emnlp-main.494/.

Ilkka J. Rissanen, Victor H. Souza, Jaakko O. Nieminen, Lari M. Koponen, and Risto J. Ilmoniemi. Advanced pipeline for designing multi-locus TMS coils with current density constraints. *IEEE Transactions on Biomedical Engineering*, 70(7):2025–2034, 2023. doi: 10.1109/TBME.2023.3234119.

Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli, and Alhussein Fawzi. Mathematical discoveries from program search with large language models. *Nature*, 625(7995):468–475, January 2024. doi: 10.1038/s41586-023-06924-6. URL https://doi.org/10.1038/s41586-023-06924-6.

Andrej Schwanke, Lyubomir Ivanov, David Salinas, Fabio Ferreira, Aaron Klein, Frank Hutter, and Arber Zela. Improving LLM-based global optimization with search space partitioning. In *International Conference on Learning Representations*, 2026. URL https://iclr.cc/virtual/2026/poster/10006592.

Thomas W. Sederberg and Scott R. Parry. Free-form deformation of solid geometric models. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '86, pp. 151–160. ACM, 1986. doi: 10.1145/15922.15903. URL https://doi.org/10.1145/15922.15903.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1): 148–175, 2016. doi: 10.1109/JPROC.2015.2494218. URL https://doi.org/10.1109/JPROC.2015.2494218.

Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Martinez Alvarado, Jacob M. Janey, Ryan P. Adams, and Abigail G. Doyle. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590(7844):89–96, 2021. doi: 10.1038/s41586-021-03213-y. URL https://doi.org/10.1038/s41586-021-03213-y.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, volume 25, pp. 2960–2968. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper_files/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html.

Enrico Spateri, Jr. Sabug, Lorenzo, Fredy Ruiz, and Giambattista Gruosso. Efficient multiobjective optimization framework for induction heating systems design. *IEEE Access*, 12:95347–95355, 2024. doi: 10.1109/ACCESS.2024.3425933.

Mohammad Hossein Tavakoli, Abdollah Ojaghi, Ebrahim Mohammadi-Manesh, and M. Mansour. Influence of coil geometry on the induction heating process in crystal growth systems. *Journal of Crystal Growth*, 311(6):1594–1599, 2009. doi: 10.1016/j.jcrysgro.2009.01.092.

R. Turner. A target field approach to optimal coil design. *Journal of Physics D: Applied Physics*, 19 (8):L147–L151, August 1986. doi: 10.1088/0022-3727/19/8/001. URL https://doi.org/10.1088/0022-3727/19/8/001.

James P. N. Uzzell, Laura R. Pickard, Ian Hamerton, and Dmitry S. Ivanov. Novel cellular coil design for improved temperature uniformity in inductive heating of carbon fibre composites. *Materials & Design*, 237:112551, January 2024. doi: 10.1016/j.matdes.2023.112551. URL https://doi.org/10.1016/j.matdes.2023.112551.

Linnan Wang, Rodrigo Fonseca, and Yuandong Tian. Learning search space partition for black-box optimization using monte carlo tree search. In *Advances in Neural Information Processing Systems*, volume 33, pp. 19511–19522. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/e2ce14e81dba66dbff9cbc35ecfdb704-Abstract.html.

Xinyuan Wang, Chenxi Li, Zhen Wang, Fan Bai, Haotian Luo, Jiayou Zhang, Nebojsa Jojic, Eric P. Xing, and Zhiting Hu. PromptAgent: Strategic planning with language models enables expert-level prompt optimization. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/686a3f32067838c8dbb68da6e9e3cf69-Abstract-Conference.html.

Ziyu Wang, Masrour Zoghi, Frank Hutter, David Matheson, and Nando de Freitas. Bayesian optimization in high dimensions via random embeddings. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1778–1784. AAAI Press, 2013. URL https://www.ijcai.org/Proceedings/13/Papers/263.pdf.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. Deep kernel learning. In Arthur Gretton and Christian C. Robert (eds.), *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pp. 370–378. PMLR, 2016. URL https://proceedings.mlr.press/v51/wilson16.html.

Wenfeng Wu, Binquan Zhou, Zhanchao Liu, Jing Wang, Haoying Pang, Linlin Chen, Wei Quan, and Gang Liu. Design of highly uniform magnetic field coils based on a particle swarm optimization algorithm. *IEEE Access*, 7:125310–125322, 2019. doi: 10.1109/ACCESS.2019.2933608.

Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. Large language models as optimizers. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/3339f19c5fcee3ad74502947a32be9e6-Abstract-Conference.html.

Haoran Ye, Jiarui Wang, Zhiguang Cao, Federico Berto, Chuanbo Hua, Haeyeon Kim, Jinkyoo Park, and Guojie Song. Reevo: Large language models as hyper-heuristics with reflective evolution. In *Advances in Neural Information Processing Systems*, volume 37, pp. 43571–43608. Curran Associates, Inc., 2024. doi: 10.52202/079017-1381. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/4ced59d480e07d290b6f29fc8798f195-Abstract-Conference.html.

Rosen Ting-Ying Yu, Cyril Picard, and Faez Ahmed. GIT-BO: High-dimensional bayesian optimization with tabular foundation models. In *International Conference on Learning Representations*, 2026. URL https://iclr.cc/virtual/2026/poster/10011090.

Xinzhe Yuan, Zhuo Chen, Jianshu Zhang, Huan Xiong, Nanyang Ye, Yuqiang Li, and Qinying Gu. Unleashing LLMs in bayesian optimization: Preference-guided framework for scientific discovery. In *International Conference on Learning Representations*, 2026. URL https://iclr.cc/virtual/2026/poster/10010010.

Mert Yuksekgonul, Federico Bianchi, Joseph Boen, Sheng Liu, Pan Lu, Zhi Huang, Carlos Guestrin, and James Zou. Optimizing generative AI by backpropagating language model feedback. *Nature*, 639(8055):609–616, 2025. doi: 10.1038/s41586-025-08661-4. URL https://doi.org/10.1038/s41586-025-08661-4.

Yimeng Zeng, Natalie Maus, Haydn Thomas Jones, Jeffrey Tao, Fangping Wan, Marcelo Der Torossian Torres, Cesar de la Fuente-Nunez, Ryan Marcus, Osbert Bastani, and Jacob R. Gardner. Scaling multi-task bayesian optimization with large language models. In *International Conference on Learning Representations*, 2026. URL https://iclr.cc/virtual/2026/poster/10008290.

Shijian Zheng, Fangxiao Jin, Shuhai Zhang, Quan Xue, and Mingkui Tan. Deep electromagnetic structure design under limited evaluation budgets. In *Proceedings of the International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 78170–78189. PMLR, 2025. URL https://proceedings.mlr.press/v267/zheng25g.html.

Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers. In *International Conference on Learning Representations*, 2023. URL https://iclr.cc/virtual/2023/poster/10850.

SEMANTIC-GUIDED MANIFOLD SEARCH FOR PHYSICAL INVERSE DESIGN

APPENDIX

CONTENTS

A Theoretical Analysis **17**
- A.1 Vectorized Free-Coordinate Representation . . . 17
- A.2 Smooth Local Branch of the Objective . . . 17
- A.3 Local Geometry of a Grounded Action . . . 17
- A.4 Interpretation of the Same-Rank Reference . . . 19

B Method Details **19**
- B.1 Semantic Action Planner . . . 20
- B.2 Semantic-Action DSL . . . 21
- B.3 Geometry Grounding Translator . . . 22
- B.4 Surrogate-Guided Executor . . . 23

C SimCoil-Bench Specification **24**
- C.1 Simulator and Shared Excitation . . . 24
- C.2 Benchmark Geometries . . . 24
- C.3 Evaluation Metrics . . . 25
- C.4 Path Feasibility and Obstacle Handling . . . 26

D Implementation Details and Experimental Protocol **27**
- D.1 Computing Environment and Randomness . . . 27
- D.2 SemGMS Configuration . . . 27
- D.3 Budget and Evaluation Protocol . . . 28
- D.4 Baseline Configuration and Shared Feasibility . . . 28
- D.5 LLM Input and Decoding Alignment . . . 28

E Additional Experimental Results **29**
- E.1 Multi-Seed Significance Analysis . . . 29
- E.2 Effect of Coil Layers . . . 30
- E.3 Comparison with Full-Space and Parametric Optimization . . . 31
- E.4 Effect of LLM Backbones Across Methods . . . 31
- E.5 Additional Ablations . . . 32
- E.6 Visualization Results . . . 32

F Limitations and Discussion **33**

G Complete LLM Prompt and Output Examples **34**
- G.1 Interface Conventions . . . 34
- G.2 Round-Level Evidence Interpreter . . . 35
- G.3 Cross-Round Experience Distiller . . . 45
- G.4 Semantic Action Planner . . . 53

A Theoretical Analysis

A.1 Vectorized Free-Coordinate Representation

Let $V_{\text{free}} = V \backslash V_{\text{fix}}$ and $D = 3|V_{\text{free}}|$. The operator vec_{free} stacks the free entries of a coil embedding into a vector in $\mathbb{R}^D$; fixed entries remain implicit. For notational consistency with the main paper, we continue to write P_t for $\text{vec}_{\text{free}}(P_t)$. Likewise, vectorizing the GGT modes $d_h^{(1)}, \dots, d_h^{(d_h)}$ gives

$$\mathbf{B}_h = \left[\text{vec}_{\text{free}}(d_h^{(1)}), \dots, \text{vec}_{\text{free}}(d_h^{(d_h)})\right] \in \mathbb{R}^{D \times d_h}. \tag{10}$$

Thus, $\mathbf{B}_h\theta$ is exactly the vectorized free-coordinate displacement induced by the action coefficients. Let $\bar{d}_h = \text{rank}(\mathbf{B}_h)$ and let $Q_h \in \mathbb{R}^{D \times \bar{d}_h}$ have orthonormal columns spanning $\text{Range}(\mathbf{B}_h)$. Then $A_h = Q_h Q_h^\top$ is the orthogonal projector onto the grounded deformation span. In particular,

$$\rho_h = \frac{\|Q_h^\top g_t\|_2^2}{\|g_t\|_2^2} = \frac{g_t^\top A_h g_t}{\|g_t\|_2^2} \in [0, 1], \tag{11}$$

because the eigenvalues of A_h are either zero or one.

A.2 Smooth Local Branch of the Objective

Consider the max-type task term used in the main paper, $\mathcal{L}_{\text{task}}(Y(P)) = \max_m \ell_m(P)$. Suppose that its active index at P_t is unique, denoted by m_t, with margin

$$\Delta_t = \ell_{m_t}(P_t) - \max_{m \neq m_t} \ell_m(P_t) > 0. \tag{12}$$

Continuity of the response terms implies the existence of a neighborhood $\mathcal{U}_t$ in which m_t remains active. If the active branches of the piecewise penalty also remain unchanged on $\mathcal{U}_t$, then the scalar objective coincides there with one branch,

$$F(P) = \ell_{m_t}(P) + \lambda \Phi_{b_t}(P), \qquad P \in \mathcal{U}_t, \tag{13}$$

where b_t denotes the active penalty branch. If the gradients of ℓ_{m_t} and Φ_{b_t} are respectively $L_{\ell,t}$- and $L_{\Phi,t}$-Lipschitz, then F is L-smooth on $\mathcal{U}_t$ with, for example, $L = L_{\ell,t} + \lambda L_{\Phi,t}$. Consequently, for any segment contained in $\mathcal{U}_t$ and any displacement s along that segment,

$$F(P_t + s) \leq F(P_t) + \langle g_t, s \rangle + \frac{L}{2}\|s\|_2^2, \qquad g_t = \nabla F(P_t). \tag{14}$$

This is the local smoothness relation used in Proposition 1.

A.3 Local Geometry of a Grounded Action

Recall that GGT assigns the coefficient box

$$\Theta_h = \left\{\theta \in \mathbb{R}^{d_h} : |\theta_j| \leq \bar{\theta}_{h,j},\ j = 1, \dots, d_h\right\}. \tag{15}$$

For grounded actions whose projection-inactive coefficient domain contains a Euclidean neighborhood of zero, let $\Theta_h^{\text{loc}} \subseteq \Theta_h$ denote such a neighborhood. On this region, the grounded map reduces to

$$\psi_h(\theta) = P_t + \mathbf{B}_h\theta, \qquad \theta \in \Theta_h^{\text{loc}}. \tag{16}$$

Its image

$$\mathcal{M}_h^{\text{loc}} = \{P_t + \mathbf{B}_h\theta : \theta \in \Theta_h^{\text{loc}}\} \subseteq \mathcal{M}_h \tag{17}$$

lies in the affine space $P_t + \text{Range}(\mathbf{B}_h)$. After redundant or zero-width modes are removed, this image is a rank-$\bar{d}_h$ affine manifold patch, with boundaries or corners inherited from the coefficient bounds. Projection, rejection, and intersections of action patches may make the complete grounded family piecewise smooth; Proposition 1 uses the local patch in Eq. (17).

We next derive the relative-ball condition in Proposition 1 from the grounded coefficient domain. Since zero lies in the interior of Θ_h^{loc}, there exists $r_{\theta,h} > 0$ such that

$$\mathbb{B}_{d_h}(0, r_{\theta,h}) = \{\theta : \|\theta\|_2 \leq r_{\theta,h}\} \subseteq \Theta_h^{\text{loc}}. \tag{18}$$

For the box in Eq. (15), one may choose any $r_{\theta,h} \le \min_j \bar{\theta}_{h,j}$ that remains inside the projection-inactive neighborhood. Let $\sigma_h^+ > 0$ be the smallest nonzero singular value of $\mathbf{B}_h$ and set

$$\nu_h = \sigma_h^+ r_{\theta,h}. \tag{19}$$

For any $s \in \mathrm{Range}(\mathbf{B}_h)$ satisfying $\|s\|_2 \le \nu_h$, choose the minimum-norm coefficient $\theta_s = \mathbf{B}_h^\dagger s$. Since $\mathbf{B}_h \mathbf{B}_h^\dagger s = s$ on $\mathrm{Range}(\mathbf{B}_h)$ and $\|\mathbf{B}_h^\dagger\|_2 = 1/\sigma_h^+$,

$$\|\theta_s\|_2 \le \|\mathbf{B}_h^\dagger\|_2 \|s\|_2 \le \frac{\nu_h}{\sigma_h^+} = r_{\theta,h}. \tag{20}$$

Hence $\theta_s \in \Theta_h^{\mathrm{loc}}$ and $P_t + s = \psi_h(\theta_s) \in \mathcal{M}_h^{\mathrm{loc}}$. Therefore,

$$\{P_t + s : s \in \mathrm{Range}(\mathbf{B}_h),\ \|s\|_2 \le \nu_h\} \subseteq \mathcal{M}_h^{\mathrm{loc}}. \tag{21}$$

This establishes the local reach condition from the GGT coefficient bounds and the local feasibility margin, without reference to the objective gradient. This is the symmetric local coefficient domain used by Proposition 1 and the main-paper formulation; zero-width modes are removed before the analysis.

Proof of Proposition 1. Let $A_h = Q_h Q_h^\top$ and define the projected descent displacement

$$s_h = -\alpha A_h g_t. \tag{22}$$

Because $\rho_h > 0$, $A_h g_t \neq 0$. The step-size condition in Proposition 1 implies

$$\|s_h\|_2 = \alpha \|A_h g_t\|_2 \le \nu_h. \tag{23}$$

Moreover, $s_h \in \mathrm{Range}(\mathbf{B}_h)$. Equation (21) therefore gives

$$P_t + s_h \in \mathcal{M}_h^{\mathrm{loc}} \subseteq \mathcal{M}_h \subseteq \mathcal{M}_t. \tag{24}$$

The relative ball is convex, so the segment from P_t to $P_t + s_h$ remains in the local patch on which Eq. (14) holds. Since $A_h = A_h^\top = A_h^2$,

$$\langle g_t, s_h \rangle = -\alpha g_t^\top A_h g_t = -\alpha \|Q_h^\top g_t\|_2^2, \tag{25}$$

$$\|s_h\|_2^2 = \alpha^2 g_t^\top A_h g_t = \alpha^2 \|Q_h^\top g_t\|_2^2. \tag{26}$$

Substituting Eqs. (25) and (26) into Eq. (14) yields

$$F(P_t + s_h) \le F(P_t) - \alpha \left(1 - \frac{L\alpha}{2}\right) \|Q_h^\top g_t\|_2^2. \tag{27}$$

Every feasible SGE query is instantiated from $\mathcal{M}_t$, and therefore $\widehat{P}_t \in \mathcal{M}_t$. By the definition of the execution gap,

$$F(\widehat{P}_t) = \inf_{P \in \mathcal{M}_t} F(P) + \varepsilon_t \le F(P_t + s_h) + \varepsilon_t. \tag{28}$$

Combining Eqs. (27) and (28), and using $\|Q_h^\top g_t\|_2^2 = \rho_h \|g_t\|_2^2$, gives

$$F(\widehat{P}_t) \le F(P_t) - \alpha \left(1 - \frac{L\alpha}{2}\right) \rho_h \|g_t\|_2^2 + \varepsilon_t, \tag{29}$$

which is the claimed bound. Since $0 < \alpha \le 1/L$, $\alpha(1 - L\alpha/2) > 0$. Consequently,

$$\varepsilon_t < \alpha \left(1 - \frac{L\alpha}{2}\right) \rho_h \|g_t\|_2^2 \quad \Longrightarrow \quad F(\widehat{P}_t) < F(P_t). \tag{30}$$

Proof of Lemma 1. Let $Q_{\mathrm{iso}} \in \mathbb{R}^{D \times \bar{d}_h}$ have orthonormal columns spanning a Haar-uniform element of the Grassmannian $\mathrm{Gr}(\bar{d}_h, D)$, sampled independently of fixed g_t. Its alignment is

$$\rho_{\mathrm{iso}} = \frac{\|Q_{\mathrm{iso}}^\top g_t\|_2^2}{\|g_t\|_2^2}. \tag{31}$$

Let $A_{\text{iso}} = Q_{\text{iso}} Q_{\text{iso}}^\top$. Haar invariance implies that, for every orthogonal matrix $R \in \mathbb{R}^{D\times D}$, $RA_{\text{iso}}R^\top$ and A_{iso} have the same distribution. Hence

$$R\,\mathbb{E}[A_{\text{iso}}]R^\top = \mathbb{E}[A_{\text{iso}}] \qquad \text{for every orthogonal } R. \tag{32}$$

The only matrices invariant under all orthogonal conjugations are scalar multiples of the identity, so $\mathbb{E}[A_{\text{iso}}] = cI_D$. Taking traces gives

$$cD = \operatorname{tr}(\mathbb{E}[A_{\text{iso}}]) = \mathbb{E}[\operatorname{tr}(A_{\text{iso}})] = \bar{d}_h, \tag{33}$$

because every rank-$\bar{d}_h$ orthogonal projector has trace $\bar{d}_h$. Therefore,

$$\mathbb{E}[A_{\text{iso}}] = \frac{\bar{d}_h}{D} I_D. \tag{34}$$

Using Eq. (31) and conditioning on fixed g_t,

$$\mathbb{E}[\rho_{\text{iso}}] = \frac{g_t^\top \mathbb{E}[A_{\text{iso}}] g_t}{\|g_t\|_2^2} = \frac{\bar{d}_h}{D}.$$

which proves Lemma 1.

A.4 Interpretation of the Same-Rank Reference

For a common admissible step size α, define the descent term associated with a rank-$\bar{d}_h$ subspace by

$$\mathsf{Desc}(\rho;\alpha) = \alpha\left(1 - \frac{L\alpha}{2}\right)\rho\|g_t\|_2^2. \tag{35}$$

Lemma 1 gives

$$\mathbb{E}[\mathsf{Desc}(\rho_{\text{iso}};\alpha)] = \alpha\left(1 - \frac{L\alpha}{2}\right)\frac{\bar{d}_h}{D}\|g_t\|_2^2. \tag{36}$$

Thus, whenever $\rho_h > \bar{d}_h/D$, the grounded action has a larger local descent term than the same-rank isotropic reference in expectation. Combined with Eq. (30), this yields the decomposition used in the main paper: SAP and GGT determine the descent information retained by the grounded family through ρ_h, while SGE determines whether the finite query set makes ε_t small enough to realize that descent.

B Method Details

This section expands the implementation of the three modules defined in the main paper. SAP internally uses three role-specialized LLM stages: a round-level evidence interpreter, a cross-round experience distiller, and a Semantic Action Planner. These stages respectively interpret one simulator transition, consolidate evidence across transitions, and decide the next semantic action. They are internal stages of SAP rather than additional method layers. GGT and SGE remain a deterministic geometric translator and a numerical optimizer, respectively.

The three LLM stages use the same backbone and decoding settings reported in the main paper, but receive different system instructions, context views, and output schemas. A representative final output is shown below, while complete normalized prompt and response examples are provided in Section G. Their information flow is

$$(Y_t, C_t, \Delta P_t, h_{t-1}) \longrightarrow e_t \longrightarrow \mathcal{E}_t \longrightarrow \delta_t, \tag{37}$$

where ΔP_t is the measured change from the previous verified coil, e_t is a single-round evidence card, $\mathcal{E}_t$ is the distilled case memory, and δ_t is the semantic-action DSL program. This staged design prevents tentative planning hypotheses from being written directly into the verified evidence memory.

Table 5: Role contracts of the three LLM stages inside SAP. Each stage receives a deliberately restricted context and produces a typed intermediate object.

LLM stage	Problem addressed	Input	Output
Evidence interpreter	Simulator records lack semantic attribution	Verified transition, responses, geometry change, and local diagnostics	Single-round evidence card e_t
Experience distiller	Raw histories are long, redundant, and state dependent	Previous memory and recent verified evidence cards	Bounded case memory $\mathcal{E}_t$
Action planner	Evidence does not specify the next search direction	Current evidence, case memory, semantic graph, and DSL schema	One semantic-action program δ_t

B.1 Semantic Action Planner

Stage 1: Round-Level Evidence Interpreter. Why it is needed. A simulator returns numerical responses, while a useful optimization record must also state which executed transition produced them and how that transition affected the physical bottleneck. Deterministic preprocessing can compute metric margins, channel rankings, and the measured geometry delta, but these heterogeneous records do not by themselves provide a concise semantic attribution. Passing them directly into long-term memory would also allow failed, restored, or schema-rejected transitions to be misattributed. The first LLM therefore converts exactly one verified transition into an evidence card. It is explicitly forbidden from proposing the next action or forming a cross-round rule.

Input and output. Its input contains the current metric snapshot, enabled diagnostic channels, the previous executed transition, the measured geometry change, a minimal path-structure view, and a weak physical coupling prior. Its output records evidence quality, objective change, the current blocker, observed effect, prior consistency, transition attribution, and short evidence notes. Numeric objective fields and blocker scores are overwritten by the deterministic records after parsing, so the LLM supplies interpretation rather than metric calculation.

Stage 2: Cross-Round Experience Distiller. Why it is needed. Even verified evidence grows linearly with the simulation budget, and the same semantic action can have different effects at different design states. Simple recency truncation may overemphasize an isolated round, whereas a flat history makes repeated successes, reversals, and no-op patterns difficult to identify. The second LLM therefore consolidates semantically related evidence across rounds. It maintains bounded memory rather than making an optimization decision: a single observation remains evidence, while only repeated verified outcomes are retained as reusable experience.

Input and output. The Distiller receives the previous case-memory index, the ten most recent verified evidence cards, metric semantics, scene values, and the shared physical prior. It returns $\mathcal{E}_t$ with five bounded fields: the best verified state, repeated observations, failed or no-op patterns, a search bias, and a recent evidence digest. It cannot emit a DSL or geometry edit. This stage realizes the cross-round compression in $\mathcal{E}_t = \varphi_{\mathrm{mem}}(\mathcal{H}_t)$. A deterministic renderer serializes the returned JSON index as compact Markdown for later prompt injection.

Stage 3: Semantic Action Planner. Why it is needed. The evidence card states what happened and the Distiller summarizes what has been reliably observed, but neither determines the next search direction. This requires jointly reasoning over the current blocker, channel-level spatial evidence, the present coil structure, prior action outcomes, and the typed action schema. The final LLM performs this decision while leaving all coordinate-level choices to GGT and SGE.

Input and output. The Planner receives the current metric and channel snapshots, the semantic graph view $G_t^{\mathrm{sem}} = \xi(G, P_t)$, $\mathcal{E}_t$, scene semantics, physical priors, exploration state, and the admissible DSL schema. It returns exactly one program $\delta_t = \pi_{\mathrm{SAP}}(\mathbf{s}_t)$. The program specifies the objective focus, semantic group, axes, qualitative direction, coarse range, and search policy. It may not contain coordinates, point indices, numerical coefficients, repair instructions, or multiple alternatives. A deterministic parser validates every field against the DSL schema and active semantic

Table 6: Fields of the semantic-action DSL. The nested fields define the geometric action, while the top-level fields retain its family and context.

Field	Type or values	Meaning
`proposal_type`	`semantic_trust_region`	Action family to be grounded by GGT
`objective_focus`	Nonempty metric list	Physical response currently prioritized
`causal_hypothesis`	Short string	Evidence-grounded rationale for the action
`search_policy`	`exploit`, `explore`, `escape_plateau`	Search posture passed to the executor
`semantic_groups`	Active graph-group identifiers	Coarse focus κ on the current coil
`axes`	Nonempty subset of $\{x, y, z\}$	Coordinate axes used by the action operator o
`direction`	Typed qualitative token	Signed or target-relative direction η
`scale`	`micro`, `small`, `medium`, `large`	Coarse action range β

groups. Invalid JSON is repaired at most once; otherwise, a conservative typed fallback is used. The main paper writes the corresponding graph-valued input compactly as G in $\mathbf{s}_t = (\mathcal{E}_t, Y_t, C_t, G)$. Here, G_t^{sem} denotes its implementation-level serialization: the edge set is inherited from G, while current node attributes are deterministically computed from P_t. This notation introduces no additional simulator signal.

B.2 Semantic-Action DSL

Purpose. Free-form instructions can express physical intent, but their scope, direction, and magnitude may be ambiguous to a numerical optimizer. The semantic-action domain-specific language (DSL) is therefore the typed interface between SAP and GGT. It preserves the LLM's high-level decision while excluding coordinates, point indices, and deformation coefficients. Consequently, every accepted program specifies where the subsequent search should be concentrated without directly constructing a coil.

Format and semantics. A program δ_t is a single JSON object with four top-level fields and a nested trust-region object. Table 6 defines the fields. The admissible metric names, semantic groups, and direction tokens are instantiated from the current benchmark and coil graph rather than fixed globally.

Semantic-Action DSL Format

```
{
  "proposal_type": "semantic_trust_region",
  "objective_focus": ["<active metric>", "..."],
  "causal_hypothesis": "<evidence-grounded rationale>",
  "search_policy": "exploit | explore | escape_plateau",
  "trust_region": {
    "semantic_groups": ["<active semantic group>", "..."],
    "axes": ["x | y | z"],
    "direction": "positive | negative | either |
                  toward_welding_region | away_from_welding_region",
    "scale": "micro | small | medium | large"
  }
}
```

Typed interpretation. The general formulation admits a typed operator library $\mathcal{O}$. All experiments reported in this paper instantiate $\mathcal{O}_{\text{exp}} = \{\text{Translate}\}$; the remaining DSL fields determine its focus, axes, direction, and range. Other operators are possible extensions but are not evaluated here. The parser first checks JSON syntax, required fields, enumerated values, benchmark-visible objectives, and graph-active group identifiers. For an accepted trust-region program, its geometric

fields are interpreted as the action

$$\begin{aligned} h &= (o, \kappa, \eta, \beta), \\ o &= \text{Translate}(\texttt{axes}), \\ \kappa &= \texttt{semantic_groups}, \\ \eta &= \texttt{direction}, \qquad \beta = \texttt{scale}. \end{aligned} \tag{38}$$

The reported implementation emits one deformation clause per round, so $\mathcal{D}_t = \tau_{\text{DSL}}(\delta_t) = (h)$ and $L_t = 1$. The remaining fields specify the optimization objective, causal rationale, and search posture; they do not add geometric degrees of freedom. This separation keeps the main-paper action tuple compact while retaining an auditable record of why the action was issued.

Semantic-Action DSL Example: Normalized Program δ_t

```
{
  "proposal_type": "semantic_trust_region",
  "objective_focus": [
    "uniformity_error"
  ],
  "causal_hypothesis": "Tube pair 2 has the lowest lower- and upper-
tube responses. Moving the upper layer rightward should increase
coupling to this relatively weaker pair and reduce inter-pair loss
imbalance.",
  "search_policy": "explore",
  "trust_region": {
    "semantic_groups": [
      "layer_2_from_bottom"
    ],
    "axes": [
      "x"
    ],
    "direction": "positive",
    "scale": "medium"
  }
}
```

This instance prioritizes loss uniformity and directs the next search toward a positive x translation of `layer_2_from_bottom` with a medium range to strengthen the relatively weaker tube pair. It does not describe a new coil. The exact graph support, deformation modes, coefficient bounds, and numerical coefficients remain unresolved until GGT interprets the program.

B.3 Geometry Grounding Translator

The semantic-action program identifies a physically motivated search direction, but its language-level focus and symbolic range cannot yet be applied to the current embedding. GGT resolves this gap without another LLM call. It parses the validated DSL into an ordered action list $\mathcal{D}_t$ and resolves each focus against the current semantic graph rather than fixed coordinate indices. Thus, the same label, such as an upper layer or a region behind the welding targets, remains meaningful after resampling or a change in layer count. The resolved group forms the core support S_h, while fixed power-interface paths are excluded.

Mode construction. The experimental implementation uses a coordinated group-displacement mode and up to three local hotspot modes, following the general use of coupled control-point deformations in geometric modeling (Sederberg & Parry, 1986). A group mode applies the same axis-aligned displacement to all points in the selected semantic group. A hotspot mode adds a local correction at a high-responsibility point in the same region. These columns instantiate the deformation basis $\mathbf{B}_h$. Because neighboring points can share a group mode, the resulting motion is coupled at the graph level rather than being independent coordinate perturbations.

Range resolution and feasibility. The symbolic ranges `micro`, `small`, `medium`, and `large` map to nominal radii of 0.5, 1.0, 2.0, and 4.0 mm, respectively. The direction η orients the dis-

placement modes, and each radius defines the symmetric coefficient interval used in the main paper. GGT probes each interval at eight locations and contracts it when a hard path condition is encountered. The checks cover self-intersection, obstacle collision, minimum clearance, path connectivity, and routing order. Joint coefficient vectors are checked again after decoding by ψ_h, so infeasible combinations are rejected before simulation.

GGT Interface for the Same Iteration

```
INPUT h
operator: coordinated translation
focus: layer_2_from_bottom
direction: +x
range: medium

OUTPUT
support S_h:
  all control points in layer_2_from_bottom
basis B_h:
  mode 1: coherent +x displacement of the upper layer
  modes 2--4: local +x adjustments at three hotspots
coefficient domain:
  Theta_h = [-2.0, 2.0]^4 mm
feasibility report:
  4 valid modes; no bound contraction in this iteration
```

For this instance, $d_h = 4$. The first coefficient controls a coherent upper layer displacement, while the remaining coefficients refine three local positions within the displacement modes oriented by SAP. GGT therefore converts one language-level action into a bounded four-variable deformation family. It neither evaluates the physical response nor selects the final coefficient vector.

B.4 Surrogate-Guided Executor

Grounding determines the admissible deformation variables, but the physical response associated with their magnitudes remains available only through the expensive simulator. SGE handles this numerical uncertainty without an LLM. It receives the grounded families and the observation set $\mathcal{I}_t = \{(P_i, F_i)\}_{i=1}^{N_t}$. The first 100 simulator evaluations initialize this model and the evidence memory. The executor models $U(P) = -F(P)$, preserving the minimization convention of the main paper while allowing acquisition maximization. We use a BoTorch `SingleTaskGP` with standardized outputs and qLogExpected Improvement (Balandat et al., 2020; Ament et al., 2023).

The surrogate input is the decoded embedding feature $\mathrm{vec}(P)$ rather than textual history. Acquisition optimization is nevertheless restricted to the current action variables: SGE samples $\theta \in \Theta_h$, forms $P = \psi_h(\theta)$, removes geometrically infeasible designs, and scores only the remaining embeddings. Thus, evaluated embeddings provide numerical supervision, whereas the grounded domains determine where the acquisition rule may query.

For each execution step, 2,048 coefficient vectors are generated across the current grounded families. Feasible designs are ranked by qLogExpected Improvement, and the highest-ranked design is sent to the simulator ($K = 1$). Its physical responses and objective are appended to $\mathcal{H}_{t+1}$ with the originating action, enabling the next round-level interpreter to attribute the outcome to the action that produced it.

SGE Interface for the Same Iteration

```
INPUT
observations: 188 evaluated embeddings and utilities
family: M_h = {psi_h(theta) | theta in [-2.0,2.0]^4}
surrogate: SingleTaskGP
acquisition: qLogExpected Improvement
coefficient pool: 2,048 feasible-screened trials

OUTPUT
theta = [1.994, 1.737, 1.781, 1.818] mm
query status: feasible
simulator result:
  welding loss: 167.43 -> 167.07
  uniformity error: 4.35% -> 3.62%
  utility: 3.65 -> 4.38
```

The complete transition therefore preserves a strict division of labor. The three LLM stages turn verified simulator outcomes into a semantic action, GGT constructs its executable variables, and SGE determines their numerical values. The new simulator result then updates both the numerical observations used by SGE and the verified semantic evidence used by SAP.

C SimCoil-Bench Specification

C.1 Simulator and Shared Excitation

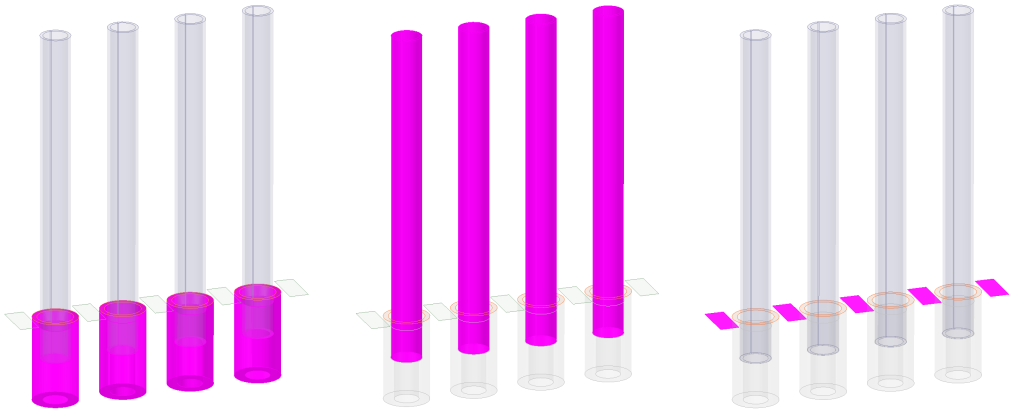
All benchmark instances are evaluated in ANSYS Maxwell 19.2. Across the three tasks, the coil conductor has a diameter of $8\,\mathrm{mm}$ and is driven by an input current of $2000\,\mathrm{A}$. These quantities are held fixed throughout optimization, so the evaluated response changes only with the coil geometry. The default scenes contain $N = 4$ parallel target-tube pairs, as illustrated in Figure 6. Absolute coordinates are instance-specific and vary with the randomly generated initialization. We therefore define each benchmark by its relative geometry and physical dimensions, while the fixed power-interface paths provide the boundary conditions. Under the notation of the main paper, the default optimizable coil body contains 11 ordered control points connected by 10 directed edges. The complete graph $G = (V, E)$ also contains the fixed inlet and outlet power-interface paths indexed by V_{fix}; their coordinates and connectivity remain unchanged.

C.2 Benchmark Geometries

Welding Loss Optimization (WLO). Each target consists of a coaxial upper–lower tube pair. The upper and lower tubes have heights of $70\,\mathrm{mm}$ and $20\,\mathrm{mm}$, respectively, and their separation is $15\,\mathrm{mm}$. WLO evaluates the induced loss over the welding region of each tube and rewards both a large worst-pair response and balanced energy delivery across the four pairs. This emphasis on response magnitude and spatial balance follows prior induction-heating coil optimization studies (Spateri et al., 2024; Choubey et al., 2025; Uzzell et al., 2024); the exact benchmark metrics are defined below.

Magnetic Field Optimization (MFO). MFO replaces the discrete welding regions with a three-dimensional target volume of length $50\,\mathrm{mm}$, width $15\,\mathrm{mm}$, and height $19\,\mathrm{mm}$. It evaluates the mean magnetic-field magnitude throughout this volume together with the pointwise field-uniformity error, reflecting the target-field and field-homogeneity objectives used in coil design (Turner, 1986; Wu et al., 2019).

Hybrid Field-Loss Optimization (HFLO). HFLO uses the same tube dimensions and $15\,\mathrm{mm}$ inter-tube separation as WLO. A $10\,\mathrm{mm} \times 4\,\mathrm{mm}$ field-evaluation plane is centered between the upper and lower tubes of each pair. These four equal-area cross sections sample the HFLO target-field region summarized in the main paper. The task jointly evaluates welding loss, loss balance, and field balance across the planes, testing whether an optimizer can allocate attention across heterogeneous physical feedback.



(a) Lower-Tube Welding Regions (b) Upper-Tube Welding Regions (c) Field-Evaluation Planes

Figure 6: Evaluation regions in SIMCOIL-BENCH: (a) Lower-Tube Welding Regions, (b) Upper-Tube Welding Regions, and (c) four $10 \times 4\,\mathrm{mm}^2$ Field-Evaluation Planes centered between the paired tubes in HFLO.

C.3 EVALUATION METRICS

Let $q_{\mathrm{EM}}(\mathbf{x})$ denote the simulated surface EM-loss density and $\mathbf{B}(\mathbf{x})$ the magnetic-field vector. For tube pair i, let Γ_i^u and Γ_i^l denote the corresponding upper- and lower-tube evaluation surfaces, respectively.

Welding loss and pairwise uniformity. We first integrate the EM-loss density over each tube surface,

$$\ell_i^r = \int_{\Gamma_i^r} q_{\mathrm{EM}}(\mathbf{x})\,\mathrm{d}A, \qquad r \in \{u, l\}. \tag{39}$$

This is the quantity returned by Maxwell's surface-integral report for the corresponding named tube surface. The reported loss conservatively sums the weakest lower- and upper-tube responses:

$$L = \min_{1\le i\le N} \ell_i^l + \min_{1\le i\le N} \ell_i^u. \tag{40}$$

This channel-wise aggregation prevents a strong response on one tube side from hiding an underserved region on the other. For each side, we compute a range-based spread and average the two spreads:

$$E_L = \frac{1}{2} \sum_{r\in\{u,l\}} \frac{\max_i \ell_i^r - \min_i \ell_i^r}{\max_i \ell_i^r}. \tag{41}$$

Thus, larger L and smaller E_L indicate stronger and more balanced welding responses. WLO reports both quantities, while HFLO retains them as its loss-related metrics.

Volumetric magnetic-field metrics. Let Ω_B be the MFO target volume. Field strength is the spatial mean of the field magnitude,

$$S_B = \frac{1}{|\Omega_B|} \int_{\Omega_B} \|\mathbf{B}(\mathbf{x})\|_2\,\mathrm{d}V. \tag{42}$$

Its pointwise uniformity error is computed from the extrema inside the same volume:

$$\begin{aligned} E_B^{\mathrm{vol}} &= \frac{B_{\max}^{\Omega} - B_{\min}^{\Omega}}{B_{\max}^{\Omega}}, \\ B_{\max}^{\Omega} &= \max_{\mathbf{x}\in\Omega_B} \|\mathbf{B}(\mathbf{x})\|_2, \\ B_{\min}^{\Omega} &= \min_{\mathbf{x}\in\Omega_B} \|\mathbf{B}(\mathbf{x})\|_2. \end{aligned} \tag{43}$$

MFO therefore favors large S_B and small E_B^{vol}.

Planar magnetic-field uniformity. For the ith HFLO evaluation plane Π_i, we compute the area-integrated field magnitude

$$\Phi_i = \int_{\Pi_i} \|\mathbf{B}(\mathbf{x})\|_2 \, \mathrm{d}A. \tag{44}$$

Because all four planes have the same area, comparing these integrals is equivalent to comparing their area-averaged fields. HFLO field-uniformity error is consequently

$$E_B^{\mathrm{plane}} = \frac{\max_i \Phi_i - \min_i \Phi_i}{\max_i \Phi_i}. \tag{45}$$

HFLO reports L, E_L, and E_B^{plane}, retaining the individual metrics so that improvement in one physical response cannot obscure degradation in another.

C.4 Path Feasibility and Obstacle Handling

Bounded action coefficients control deformation magnitude, but do not alone prevent nonadjacent coil segments from intersecting or the finite-diameter conductor from entering environmental geometry. We therefore instantiate the hard constraints in $\mathcal{P}$ through the deterministic feasibility map $\Pi_{\mathcal{P}}$ before any simulator call.

For a directed edge $e = (i, j) \in E$, define its centerline segment and occupied capsule as

$$\begin{aligned} c_e(s; P) &= (1 - s)p_i + sp_j, \quad s \in [0, 1], \\ \mathcal{K}_e(P) &= c_e([0, 1]; P) \oplus \mathbb{B}_{r_c}. \end{aligned} \tag{46}$$

where $r_c = 4\,\mathrm{mm}$ is the conductor radius and $\oplus$ denotes the Minkowski sum. For two nonadjacent edges e and e', their centerline distance is

$$d_{e,e'}(P) = \min_{s,t\in[0,1]} \|c_e(s; P) - c_{e'}(t; P)\|_2 \,. \tag{47}$$

Self-intersection is excluded by $d_{e,e'}(P) \geq 2r_c$. Let $\mathcal{O} = \{O_k\}_{k=1}^{K_o}$ collect the target hardware, bounding boxes, base plane, and any additional avoidance regions. Environmental clearance is enforced by

$$d_{e,k}(P) = \min_{s\in[0,1],\, y\in O_k} \|c_e(s; P) - y\|_2 \geq r_c. \tag{48}$$

For a planar obstacle $O_k = \{x : n_k^\top x = b_k\}$ with admissible side $\sigma_k \in \{-1, 1\}$, the same condition is evaluated as $\min_{s\in[0,1]} \sigma_k(n_k^\top c_e(s; P) - b_k) \geq r_c$. Together with fixed interface points, these tests instantiate

$$\begin{aligned} \mathcal{P} = \{P : {} & p_v = \bar{p}_v, \ v \in V_{\mathrm{fix}}, \\ & d_{e,e'}(P) \geq 2r_c, \quad d_{e,k}(P) \geq r_c\}. \end{aligned} \tag{49}$$

Because E and its directed order are never modified, connectivity and one-way routing order are preserved by construction; only the embedding of the optimizable vertices changes.

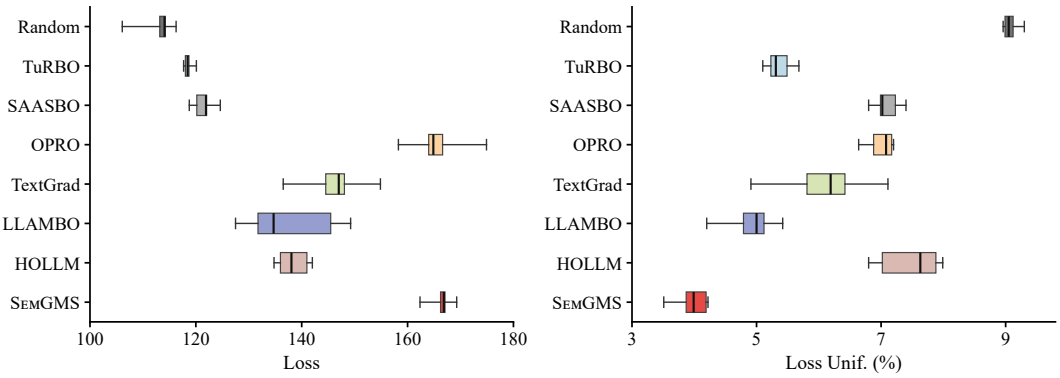
For a raw grounded proposal $\widetilde{P} = P_t + \mathbf{B}_h\theta$, we search along its displacement $\Delta P = \widetilde{P} - P_t$. The largest admissible fraction is

$$\lambda^* = \sup\{\lambda \in [0, 1] : P_t + s\Delta P \in \mathcal{P}, \ \forall s \in [0, \lambda]\}. \tag{50}$$

Operationally, eight evenly spaced probes first bracket the feasible interval, after which bisection refines its boundary. The feasibility map is

$$\Pi_{\mathcal{P}}(\widetilde{P}) = \begin{cases} \widetilde{P}, & \widetilde{P} \in \mathcal{P}, \\ P_t + 0.95\lambda^*\Delta P, & \|\lambda^*\Delta P\|_\infty \geq \epsilon_{\mathrm{move}}, \\ \perp, & \text{otherwise}, \end{cases} \tag{51}$$

where $\epsilon_{\mathrm{move}} = 10^{-6}\,\mathrm{mm}$ and $\perp$ denotes a rejected proposal. The implementation applies this clipping in directed path order and then repeats all capsule–capsule and capsule–environment tests on the joint deformation. Any residual violation is rejected before acquisition ranking or simulation. Thus, the feasibility map is a conservative backtracking operator rather than an exact Euclidean projection.



(a) Welding loss ↑ **(b) Loss uniformity error ↓**

Figure 7: Five-seed WLO distributions under 200 simulator evaluations. Boxes show the interquartile range and median, while whiskers span all five runs. Panels report (a) welding loss and (b) loss-uniformity error.

D Implementation Details and Experimental Protocol

This section records the implementation choices needed to reproduce the reported comparisons and separates method-specific optimization decisions from the common simulator and feasibility interface.

D.1 Computing Environment and Randomness

All physics evaluations use ANSYS Maxwell 19.2. Numerical optimization and LLM orchestration run on a workstation equipped with an NVIDIA RTX 3090 Ti GPU with 24 GB memory. The main comparison uses five random seeds; auxiliary experiments use seed 0 unless otherwise stated. A simulator call has a 600-second timeout; failed calls return no physical objective and are not treated as valid observations.

D.2 SemGMS Configuration

By default, SemGMS uses DeepSeek-V4-Flash with temperature 0.7, top-p 1.0, and a maximum output length of 8000 tokens. The LLM client has a 600-second timeout and retries a failed request at most twice. Initialization uses 100 simulator evaluations. Each proposal independently perturbs every optimizable coordinate by a uniform draw from $[-9.0, 9.0]\,\mathrm{mm}$ around the initial coil; duplicate or infeasible proposals are resampled, and only finite simulator outcomes enter the dataset. The resampling guard uses the configured pool parameter 64. The semantic ranges `micro`, `small`, `medium`, and `large` correspond to 0.5, 1.0, 2.0, and $4.0\,\mathrm{mm}$, respectively.

GGT constructs three hotspot modes, uses eight feasibility probes, and rejects movements below $10^{-6}\,\mathrm{mm}$. In accordance with the main paper, SGE uses a BoTorch `SingleTaskGP` with standardized outputs and qLogExpected Improvement over $U(P) = -F(P)$. Each round scores 2048 proposals using 256 Monte Carlo samples and submits one design to the simulator. The semantic loop runs for at most 100 rounds. Memory uses the ten most recent rounds, two distillation retries, and a 6000-character experience cap. The verified-round and memory-distillation contexts are capped at 20000 and 28000 characters. DSL repair is attempted once, and the no-improvement threshold is four rounds. If no feasible design exists in an SGE pool, the current anchor is retained, and the fallback is logged. With $B_{\mathrm{init}} = 0$, the cached anchor is the first numerical observation; SGE follows the semantic direction prior until at least two observations are available for surrogate fitting.

Table 7: Baseline configurations. $D = 3n$ is the flattened coordinate dimension.

Method	Initialization	Update configuration	Invalid proposal
Random	Initial coil	Independent coordinate perturbations in $[-9, 9]$ mm	Reject and resample
TuRBO	$2D$ samples	Batch 1; pool $\min(100D, 5000)$; initial trust-region length 0.8	Reject and resample
SAASBO	$2D$ samples	Batch 1; SAAS GP with EI; pool $\min(100D, 5000)$	Reject and resample
OPRO	Initial coil	Eight coordinate proposals per prompt; one retry	Line repair, then reject
TextGrad	Initial coil	One textual-gradient update per round; one retry	Line repair, then reject
LLAMBO	Five samples	Eight candidates; LLM surrogate ranks one query; one retry	Line repair, then reject
HOLLM	Eight samples	Five regions, five proposals per region, top four evaluated	Line repair, then reject

D.3 Budget and Evaluation Protocol

For each benchmark and random seed, the main comparison allocates every optimizer 200 simulator evaluations from the same initial coil; initial coils differ across seeds. For SEMGMS, the first 100 evaluations initialize the numerical surrogate and evidence memory, and the remaining 100 form the semantic-guided loop. Each SGE round ranks 2048 feasible-screened proposals and sends one design to Maxwell. The 250-evaluation experiment in the main paper changes only the evaluation budget and is reported separately as a budget-extension analysis.

D.4 Baseline Configuration and Shared Feasibility

Each baseline follows its official update rule with the adapter configuration summarized below: Random Search (Bergstra & Bengio, 2012), TuRBO (Eriksson et al., 2019), SAASBO (Eriksson & Jankowiak, 2021), OPRO (Yang et al., 2024), TextGrad (Yuksekgonul et al., 2025), LLAMBO (Liu et al., 2024b), and HOLLM (Schwanke et al., 2026).

Within each benchmark and seed, all methods receive the same initial coil, objective scalarization, simulator-output schema, and evaluation budget.

All methods use the same feasibility predicates in Section C.4. Numerical proposals from Random Search, TuRBO, and SAASBO are rejected directly when infeasible. Direct-coordinate LLM proposals are first clipped along their proposed displacement and are rejected if any hard violation remains. The method then follows the retry or attempt guard in Table 7. Only feasible coils are submitted to the expensive simulator.

D.5 LLM Input and Decoding Alignment

The first LLM-facing stage of SEMGMS is the Round-Level Evidence Interpreter, whose Verified Round Analyst prompt analyzes the current coil and its latest simulator response. We use the task preamble and initial evidence block of this prompt as the common domain prefix for all LLM-based baselines. This shared prefix specifies the benchmark objectives and metric directions, summarizes the current coil, reports the available simulator responses and feasibility status, and states the remaining evaluation budget. It contains no SEMGMS-specific DSL, grounded deformation variables, surrogate scores, or distilled memory.

After this common prefix, each baseline retains the instructions, reasoning loop, and output schema prescribed by its original method. All LLM-based methods use DeepSeek-V4-Flash and the same 8000-token output cap; method-specific decoding settings otherwise follow their official procedures. The three SEMGMS LLM stages use temperature 0.7 and top-p 1.0. This protocol equalizes domain information without rewriting the algorithmic mechanism of a baseline. Complete SEMGMS prompts and outputs are provided in Section G.

Table 8: WLO significance analysis for the five-seed results in Table 1. Means and standard deviations are repeated for reference. p_{H} is the Holm-adjusted one-sided exact Mann–Whitney U p-value against SEMGMS. Positive Cliff's δ favors SEMGMS.

Method	**Loss** ↑	p_{H}	δ	**Loss Unif. (%)** ↓	p_{H}	δ
Random	112.77±3.90	0.028	1.00	9.08±0.14	0.028	1.00
TuRBO (NeurIPS'19)	118.59±0.93	0.028	1.00	5.36±0.23	0.028	1.00
SAASBO (UAI'21)	121.50±2.20	0.028	1.00	7.09±0.23	0.028	1.00
OPRO (ICLR'24)	165.73±6.01	0.274	0.28	6.99±0.23	0.028	1.00
TextGrad (Nature'25)	146.20±6.63	0.028	1.00	6.09±0.81	0.028	1.00
LLAMBO (ICLR'24)	137.73±9.25	0.028	1.00	4.91±0.46	0.028	0.92
HOLLM (ICLR'26)	138.36±3.12	0.028	1.00	7.46±0.53	0.028	1.00
SEMGMS (Ours)	**166.36**±2.52	–	–	**3.96**±0.29	–	–

E Additional Experimental Results

E.1 Multi-Seed Significance Analysis

We provide statistical analyses of the five-seed WLO and MFO results summarized in Table 1, followed by a descriptive summary of the HFLO runs.

We evaluate all eight methods over five random seeds with different initial designs; within each seed, all methods share the initial design and the 200-evaluation budget.

Figure 7 shows the WLO distributions, and Table 8 supplements the main-table means and standard deviations with significance tests and effect sizes.

Statistical protocol. Because normality cannot be assessed reliably from five runs, we use non-parametric tests. For each metric, a Kruskal–Wallis test first evaluates the omnibus difference among all methods (Kruskal & Wallis, 1952). After rejecting the omnibus null, we compare SEMGMS with each baseline using an exact one-sided Mann–Whitney U test (Mann & Whitney, 1947), with the alternative fixed by the metric direction before comparison: higher for Loss and lower for Loss Unif. The seven comparisons within each metric are corrected by Holm's procedure (Holm, 1979). We also report Cliff's δ (Cliff, 1993), oriented so that positive values favor SEMGMS.

The omnibus differences are significant for both Loss ($H(7) = 36.66$, $p = 5.44\times10^{-6}$, $\epsilon^2 = 0.927$) and Loss Unif. ($H(7) = 34.65$, $p = 1.30 \times 10^{-5}$, $\epsilon^2 = 0.864$), with large effect sizes. SEMGMS obtains the highest mean Loss of 166.36 ± 2.52 and the lowest mean Loss Unif. of $3.96 \pm 0.29\%$. Its Loss is significantly higher than that of every baseline except OPRO after Holm correction. Against OPRO, the mean difference is small and not significant ($p_{\mathrm{H}} = 0.274$, $\delta = 0.28$). In contrast, SEMGMS significantly improves Loss Unif. over all seven baselines ($p_{\mathrm{H}} = 0.028$), with $\delta = 0.92$ against LLAMBO and $\delta = 1.00$ against the remaining methods. The result therefore supports a consistent uniformity advantage while showing that the strongest LLM baseline can attain statistically comparable welding loss.

These findings also provide direct evidence that LLM-guided optimization can transfer to spatial coil design. Prior work supplies complementary evidence for the underlying capability: OPRO uses value-annotated histories across heterogeneous optimization tasks (Yang et al., 2024), LLAMBO uses contextual and historical information for model-based black-box optimization across diverse benchmarks (Liu et al., 2024b), and TextGrad applies language feedback to scientific objectives including molecule design and radiotherapy planning (Yuksekgonul et al., 2025). Those studies do not establish coil performance, but support the premise that LLMs can convert iterative feedback into useful search signals. The five-seed WLO results provide the corresponding domain-specific evidence while also showing that structured semantic guidance is important for balancing physical responses.

MFO stability. The five-seed MFO results in Table 1 follow the same 200-evaluation protocol.

Table 9: MFO significance analysis for the five-seed results in Table 1. Means and standard deviations are repeated for reference. Statistical quantities follow Table 8; positive Cliff's δ favors SEMGMS.

Method	**Field Str.** ↑	p_{H}	δ	**Field Unif. (%)** ↓	p_{H}	δ
Random	96.44±2.74	0.028	1.00	31.74±4.06	0.028	1.00
TuRBO (NeurIPS'19)	99.28±1.35	0.040	0.92	24.71±1.15	0.151	0.60
SAASBO (UAI'21)	**105.90**±2.31	0.845	-0.36	25.60±2.55	0.155	0.44
OPRO (ICLR'24)	98.95±2.11	0.028	1.00	25.94±1.76	0.083	0.80
TextGrad (Nature'25)	102.94±1.98	0.421	0.36	27.49±2.70	0.032	0.92
LLAMBO (ICLR'24)	101.21±1.41	0.111	0.76	29.09±1.95	0.028	1.00
HOLLM (ICLR'26)	102.59±1.62	0.226	0.60	27.93±2.14	0.028	1.00
SEMGMS (Ours)	104.32±2.23	–	–	**23.45**±1.07	–	–

Table 10: WLO performance across coil layer counts under the same simulator budget. Loss-uniformity errors are reported in percent. Best results within each layer are bold.

	1 layer ($n = 4$)		**2 layers** ($n = 11$)		**3 layers** ($n = 17$)	
Method	Loss ↑	Loss Unif. ↓	Loss ↑	Loss Unif. ↓	Loss ↑	Loss Unif. ↓
Initial Design	50.09	18.26	112.43	17.44	184.37	24.65
Random	66.44	7.99	115.47	9.30	197.99	22.33
SAASBO (UAI'21)	71.38	5.44	118.76	7.40	195.42	20.93
TextGrad (Nature'25)	80.90	4.56	147.00	7.11	227.57	10.74
HOLLM (ICLR'26)	78.49	4.76	134.77	7.63	216.56	11.62
SEMGMS (Ours)	**90.17**	**3.13**	**167.07**	**3.62**	**240.76**	**8.59**

The initial design varies across seeds, while all methods share the same initial design within each seed.

Table 9 provides the corresponding significance tests and effect sizes.

The omnibus differences are significant for Field Str. ($H(7) = 28.76$, $p = 1.60 \times 10^{-4}$) and Field Unif. ($H(7) = 23.16$, $p = 1.60 \times 10^{-3}$). SEMGMS obtains the lowest mean Field Unif. of $23.45 \pm 1.07\%$ while maintaining the second-highest mean Field Str. of 104.32 ± 2.23. Its uniformity advantage is significant over Random, TextGrad, LLAMBO, and HOLLM after Holm correction, while the differences from TuRBO, SAASBO, and OPRO are not significant with five runs. For Field Str., SAASBO remains strongest on average. These results support the main-paper conclusion that SEMGMS consistently favors spatial field uniformity without substantially weakening response magnitude.

HFLO stability. Table 1 reports means and standard deviations over five HFLO runs; standard deviations are computed with the $n - 1$ denominator. SEMGMS attains the lowest loss-uniformity error in all five runs and the lowest mean field-uniformity error, while its mean welding loss is close to TuRBO's (171.00 versus 170.84).

E.2 Effect of Coil Layers

We evaluate how geometric complexity affects WLO by varying the coil from one to three layers while retaining the shared 200-evaluation protocol. The three representations contain $n = 4$, 11, and 17 optimizable control points, respectively. Because changing the number of layers changes conductor coverage and the scale of attainable loss, Loss should be compared across methods within each layer rather than directly across layers.

SEMGMS remains strongest across all layer counts. As shown in Table 10, it achieves both the highest Loss and the lowest Loss Unif. for one-, two-, and three-layer coils. Relative to the strongest

Table 11: Comparison with full-space GP and four-dimensional parametric optimization on WLO under the default seed-0, 200-evaluation protocol. Loss-uniformity errors are reported in percent.

Method	Loss ↑	Loss Unif. ↓
Full-Space GP	131.11	5.59
Parametric BO	143.10	8.19
SEMGMS (Ours)	**167.07**	**3.62**

included baseline in each setting, its Loss improves by 11.46%, 13.65%, and 5.80%, while its Loss Unif. decreases by 31.36%, 49.09%, and 20.02%, respectively.

Additional layers increase response capacity but make balance harder. The attainable Loss rises substantially with layer count for every method, whereas optimized uniformity errors generally increase as the representation grows under the fixed budget. Despite the larger search space and stronger cross-layer coupling at three layers, SEMGMS retains the best balance between response magnitude and uniformity. This behavior is consistent with semantic actions focusing the added degrees of freedom on physically relevant regions before numerical refinement.

E.3 Comparison with Full-Space and Parametric Optimization

To test whether the numerical optimizer alone explains the improvement, we compare SEMGMS with Full-Space GP, which uses the same BoTorch `SingleTaskGP` and qLogExpected Improvement acquisition function. Both methods share the default seed-0 initial coil, initialization observations, and 200-evaluation budget. Full-Space GP searches all free coil coordinates without action-conditioned deformation families, while retaining the fixed interface coordinates and geometric feasibility checks.

For regular coils, conventional optimization commonly preserves a predefined shape family and varies a small set of global winding-distribution parameters (Spateri et al., 2024). This representation is compact and well suited to numerical black-box optimization (Snoek et al., 2012), but it couples distant parts of the coil through the same global variables. We therefore construct a parametric BO baseline for WLO using four variables: coil length, coil width, upper-layer height, and lower-layer height. The topology and regular two-layer layout are fixed, and BO optimizes these variables under the same objective, feasibility checks, initialization, and 200-evaluation budget as the default single-seed comparison.

Compared with Full-Space GP, SEMGMS increases Loss from 131.11 to 167.07 (27.43%) and reduces Loss Unif. from 5.59% to 3.62% (35.24%) in this single-seed comparison. With the GP and acquisition function held fixed, this result supports the benefit of action-conditioned search-space construction under the tested evaluation budget.

Table 11 shows that parametric BO improves the regular coil within its compact design family, but remains limited by its globally coupled geometry. In contrast, SEMGMS attains 16.75% higher Loss and 55.80% lower Loss Unif. by adapting local path regions while retaining coordinated numerical refinement.

E.4 Effect of LLM Backbones Across Methods

We evaluate all LLM-based optimizers on WLO with DeepSeek-V4-Flash, Qwen3.6-Plus, and GLM-5.2. For each method, only the LLM backbone is replaced; its optimization procedure, the shared initial coil, and the 200-evaluation protocol remain unchanged. The DeepSeek-V4-Flash results reproduce the main comparison.

Table 12 shows that backbone choice changes the balance achieved by every LLM-based optimizer. Nevertheless, SEMGMS obtains the lowest Loss Unif. with all three backbones and the highest Loss with DeepSeek-V4-Flash and GLM-5.2. Under Qwen3.6-Plus, OPRO attains slightly higher Loss, whereas SEMGMS reduces its Loss Unif. from 8.17% to 3.16%. The consistent joint performance indicates that the benefit of semantic grounding and numerical execution is not specific to one LLM.

Table 12: Effect of the LLM backbone across LLM-based optimizers on WLO. Only the backbone is changed, while the remaining method-specific procedures and common evaluation protocol are fixed. Best results under each backbone are bold. Loss-uniformity errors are reported in percent.

	DeepSeek-V4-Flash		Qwen3.6-Plus		GLM-5.2	
Method	Loss ↑	Loss Unif. ↓	Loss ↑	Loss Unif. ↓	Loss ↑	Loss Unif. ↓
OPRO (ICLR'24)	160.32	6.46	**154.36**	8.17	158.49	6.01
TextGrad (Nature'25)	147.00	7.11	150.34	7.87	162.47	6.99
LLAMBO (ICLR'24)	149.24	4.83	142.39	5.40	152.11	7.30
HOLLM (ICLR'26)	134.77	7.63	128.63	7.77	132.58	8.26
SEMGMS (Ours)	**167.07**	**3.62**	150.80	**3.16**	**169.97**	**4.54**

Table 13: Ablation of the surrogate-initialization budget on WLO. The total simulator budget is fixed at 200 evaluations.

Init. Eval.	**Search Eval.**	Loss ↑	Loss Unif. ↓
0	200	160.48	4.57
50	150	152.34	4.19
100	**100**	**167.07**	**3.62**
150	50	131.03	5.16

E.5 Additional Ablations

Surrogate-initialization budget. We vary the number B_{init} of simulator evaluations reserved for initializing the surrogate and evidence memory while fixing the total WLO budget at 200. The remaining $200 - B_{\text{init}}$ evaluations are used for semantic-guided optimization; $B_{\text{init}} = 0$ denotes no dedicated initialization phase. All other settings are fixed. This sweep reallocates the full 200-evaluation budget. It is distinct from the main-paper "w/o Surrogate Initialization" variant, which retains the default 100/100 schedule but withholds the first 100 observations from SGE fitting.

Table 13 shows that the balanced 100/100 split achieves both the highest Loss and the lowest Loss Unif. With fewer initial observations, the surrogate has a weaker numerical prior; with 150 initialization evaluations, only 50 queries remain for semantic-guided refinement. The result supports the default split as a balance between model estimation and action-guided search.

Surrogate backend. We next replace only the surrogate model used by SGE, retaining the default 100/100 budget split, acquisition procedure, SAP, and GGT. We compare the default Gaussian process (GP) with Bayesian linear regression (BLR) and deep kernel learning (DKL) (Wilson et al., 2016).

As shown in Table 14, GP provides the strongest joint result under the limited budget. DKL remains competitive in Loss but has a higher uniformity error, whereas BLR degrades both metrics. This pattern is consistent with GP offering a favorable data-efficiency and flexibility trade-off for the bounded nonlinear response surfaces encountered by SGE.

E.6 Visualization Results

WLO geometry visualization. On the WLO benchmark, Figure 8 compares the common initial coil with the best feasible design returned by each method. The initial coil is derived from a canonical two-layer, four-sided encircling template that fixes only the path topology and scene scale. Using seed 0, every coordinate of each optimizable control point is perturbed once according to

$$\begin{aligned}(P_0)_{v,c} &= (P_{\text{tpl}})_{v,c} + \epsilon_{v,c},\\ \epsilon_{v,c} &\sim \mathcal{U}(-9, 9)\ \text{mm}, \quad c \in \{x, y, z\}.\end{aligned} \tag{52}$$

The resulting realization satisfies the shared feasibility checks and is held fixed across all methods; the power-interface paths remain unchanged. This protocol preserves a common routing topology while removing the exact symmetry of the template, so visual differences arise from optimization rather than method-specific initialization.

Table 14: Ablation of the SGE surrogate backend on WLO under the shared 200-evaluation protocol.

Surrogate	Loss ↑	Loss Unif. ↓
BLR	129.66	5.90
DKL	161.10	4.56
GP (default)	**167.07**	**3.62**

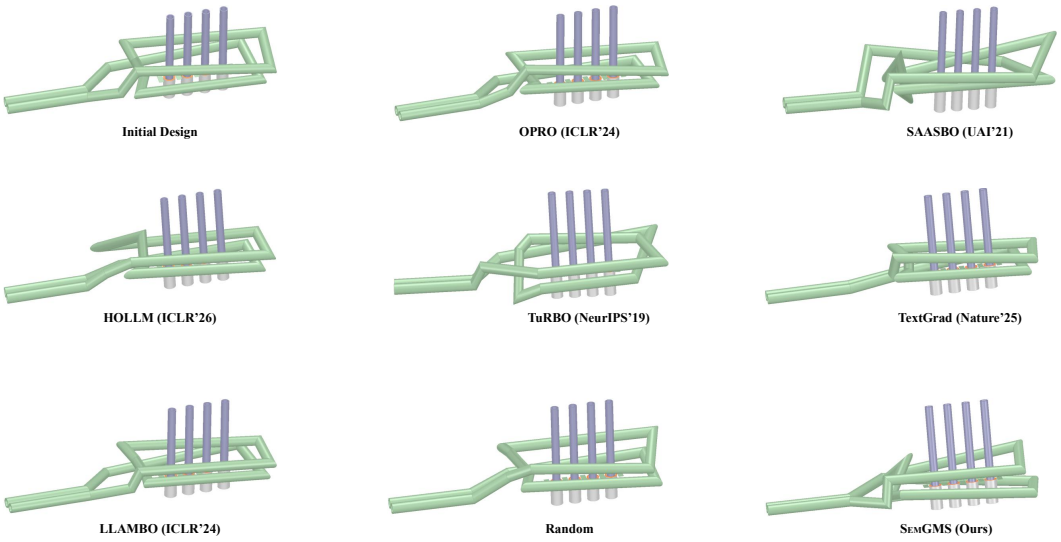


Figure 8: Qualitative WLO comparison under the shared 200-evaluation protocol. *Initial Design* is the common feasible seed-0 perturbation of the canonical two-layer, four-sided encircling template. The remaining panels show the best feasible designs returned by the compared methods.

The optimized geometries reveal different search behaviors that scalar metrics alone do not show. Several baselines make isolated or globally irregular changes, whereas SEMGMS produces coordinated local deformations while preserving the connected coil path. This behavior is consistent with its separation of semantic direction, geometry grounding, and numerical refinement; the quantitative effect is reported by the WLO results above.

Cross-objective field visualization. We next examine whether the same WLO-optimized coils retain favorable magnetic behavior under an objective not used to produce these designs. Without further optimization, we re-evaluate every coil in Figure 8 using the HFLO field protocol. A Field-Evaluation Plane is placed midway between the upper and lower tubes of each pair, and the field-strength distribution is computed over all four planes. This post-hoc test measures cross-objective transfer rather than generalization to a new geometry or scene.

As shown in Figure 9, SEMGMS retains strong field intensity while producing a visibly more uniform distribution across the evaluation planes. The result suggests that its WLO improvements do not arise from concentrating the response into a narrow region; instead, the resulting geometry remains compatible with the additional spatial-field criterion used by HFLO.

F LIMITATIONS AND DISCUSSION

The current study is limited to Maxwell-based spatial coil tasks and a fixed axis-aligned translation schema with finitely many grounded modes. The LLM also introduces inference latency and sensitivity to model updates, while the method still requires costly simulator evaluations. Extending the DSL to broader deformation operators and validating transfer across simulators and physical systems remain important directions.

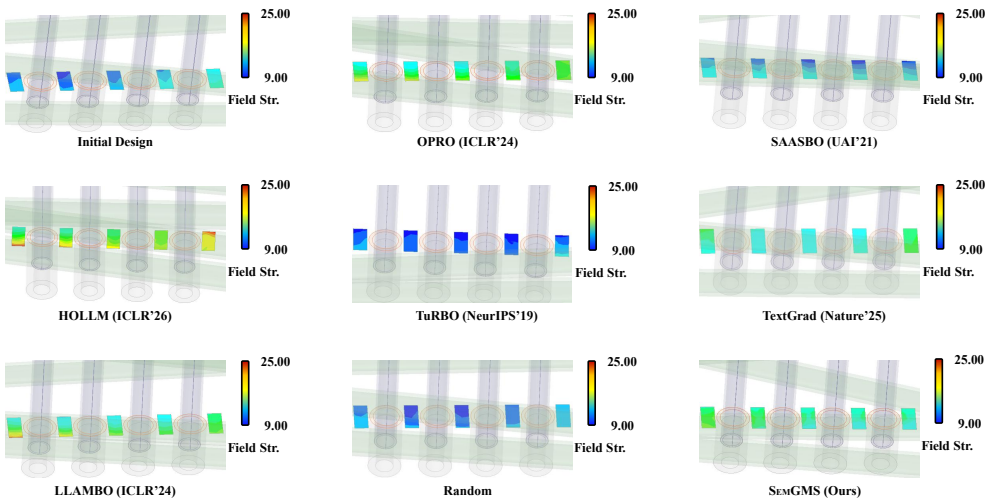


Figure 9: Post-hoc field evaluation of the same WLO designs shown in Figure 8. Following the HFLO protocol, four Field-Evaluation Planes are centered between the paired tubes, and all panels use the same field-strength scale. No method is reoptimized for this evaluation.

G Complete LLM Prompt and Output Examples

Throughout the prompt examples, the internal field `objective_value` denotes the maximized utility $U(P) = -F(P)$, rather than the minimized objective $F(P)$ used in the main paper.

All prompts use the decoding configuration in Section D.5, including the 8000-token output cap. The examples preserve the interface structure and numerical simulator records, but normalize legacy labels, metric attribution, and explanatory text to match the main paper. Context strings containing `[truncated]` or `...` were shortened by the runtime context compactor before the LLM call.

G.1 Interface Conventions

The examples correspond to verified FEM round 99. We use `verified_round_id` for the evaluated coil state and `source_iteration` for the outer-loop decision that produced or selected it. Hence, the Evidence Card uses `verified_round_id=99` and `source_iteration=98`; both are integers.

An incumbent restoration is a control-flow transition rather than a semantic action. It is therefore recorded as `transition_type=incumbent_restore`, with `dsl_validation_status=not_applicable` and a null DSL summary. The Distiller returns a structured JSON memory index; a deterministic renderer converts this index to compact Markdown for prompt injection. The parser-facing semantic-group enumeration is generated from, and therefore matches, the active groups listed in the same Planner input.

Throughout these examples, `lower_tube` and `upper_tube` refer to the Lower-Tube and Upper-Tube Welding Regions in Figure 6; `field_plane` refers to a Field-Evaluation Plane. The internal keys `weld_min_loss` and `uniformity_error` correspond to Loss and Loss Unif., respectively. The `Current Coil Structure` block is the implementation-level serialization G_t^{sem} of the graph input written as G in the main paper. Point indices occur only in this input metadata, while the Planner outputs semantic groups rather than coordinates or indices.

In the reported experiments, `semantic_trust_region` is grounded as $\mathrm{Translate}(\texttt{axes})$. The `direction` token orients the displacement modes and records SAP's preferred sign. Consistent with the symmetric coefficient domain Θ_h in the main paper, it is not a hard half-space constraint:

SGE may use either coefficient sign for local numerical refinement. The hard candidate-family restriction is imposed by the operator, focus, active axes, and bounded range. Long lines below are wrapped only for typesetting.

G.2 Round-Level Evidence Interpreter

System Prompt: Verified Round Analyst

```
You analyze a verified scientific optimization round and return
structured JSON only.
```

Prompt: Verified Round Analyst

```
You are the Verified Round Analyst. Create one evidence_card for the current verified FEM
round
only. Coil points are ordered connected 3D path, not an unordered point cloud; moving one
point
changes adjacent segments. Trust the Terminology Brief and structured evidence over prose.
Do not
invent tube-pair or proxy-point winners without channel evidence. If dsl_validation_status
is
schema_rejected, the raw semantic-action DSL was not executed; attribute the current
geometry only to the
selected DSL/SGE outcome recorded in the transition. Return JSON only with
verified_round_id, source_iteration,
evidence_quality, objective_value, previous_objective_value, objective_delta,
improvement_status,
primary_blocker, observed_effect, prior_consistency, transition_attribution, and
evidence_notes. Determine primary_blocker from the active worst normalized metric in the
scalar objective, rather than from a diagnostic channel minimum alone. Do not recommend the
 next search direction. Do not propose the next DSL, do not maintain long-term
memory, and do not promote stable lessons from one round.

## Terminology Brief
{"row_format":"term|meaning|goal_or_direction|how_to_use","terms":["coil_path|ordered
connected 3D
path; adjacent points form physical coil segments|geometry state|reason about a path, not
an
unordered point cloud","objective_value|maximized utility U(P)=-F(P) computed from verified
 FEM metrics|higher_is_better|compare verified rounds; larger value is better","
weld_min_loss|min(lower_tube_loss_values) +
min(upper_tube_loss_values)|higher_is_better|track min loss","uniformity_error|average
range-based spread of lower- and upper-tube loss channels|lower_is_better|track inter-pair
loss balance",
"lower_tube_loss_values|4 lower-tube losses by tube_pair_index|feeds
lower_tube_loss_uniformity|name weak lower region","upper_tube_loss_values|4 upper-tube
losses by
tube_pair_index|feeds upper_tube_loss_uniformity|name weak upper region","tube_pair_index
|1..4
tube-pair id|diagnostic index only|map channel evidence","weld_pair_region|semantic welding
 target
x-region between adjacent B proxy targets|true welding target region|map tube-pair loss
channels to
spatial x spans between proxy points","selection_branch|safe next-coil selection result|
process
state|accepted vs fallback","verified_round|one coil state that has completed real FEM
simulation|evidence unit|compare objective/metrics only after real simulation",
"evidence_card|compact JSON summary of one verified_round|single-round evidence|feed case
memory
and optimizer"]}

## Metric Snapshot
{"row_format":"metric|current_value|threshold|direction|display_name|margin_to_threshold",
"metrics":["weld_min_loss|167.4332528246656|115.0|higher_is_better|Welding Loss
|52.4332528246656","uniformity_error|4.351902442438149|8.0|lower_is_better|Loss Uniformity
Error|3.648097557561851"]}

## Verified FEM Evidence And Previous Change
{
  "record_type": "verified_round_record_draft",
  "verified_round_id": 99,
  "source_iteration": 98,
  "objective": {
    "objective_value": 3.65,
    "previous_objective_value": -1.69,
    "objective_delta": 5.34,
```

Prompt: Verified Round Analyst *(continued)*


```
    "improvement_status": "improved",
    "parse_status": "ok",
    "failure_mode": "none",
    "failure_interpretation": null
  },
  "primary_blocker": "uniformity_error",
  "metric_rows": [
    {
      "metric": "weld_min_loss",
      "value": 167.433253,
      "threshold": 115.0,
      "direction": "higher_is_better",
      "margin_to_threshold": 52.433253,
      "status": "meets_threshold",
      "blocker_score": 0.0
    },
    {
      "metric": "uniformity_error",
      "value": 4.351902,
      "threshold": 8.0,
      "direction": "lower_is_better",
      "margin_to_threshold": 3.648098,
      "status": "meets_threshold",
      "blocker_score": 0.0
    }
  ],
  "channel_evidence": {
    "optimization_mode": "uniformity_rebalancing",
    "actionable_channel_focus": "Uniformity error is the active worst-metric term. Tube
pair 2 has the minimum lower- and upper-tube losses, but all pairwise loss responses remain
 above the loss threshold.",
    "global_loss_deficit_summary": {
      "weld_min_loss_threshold": 115.0,
      "weld_min_loss_from_channel_mins": 167.4332528246656,
      "pair_count": 4,
      "pair_count_below_threshold": 0,
      "all_pairs_below_threshold": false,
      "largest_pair_deficit": -52.433253,
      "smallest_pair_deficit": -59.140237,
      "pair_deficit_row_format": "pair_index|pair_loss_sum|deficit_to_weld_min_threshold|
is_current_
floor",
      "pair_deficit_rows": [
        "1|170.527181|-55.527181|False",
        "2|167.433253|-52.433253|True",
        "3|172.234665|-57.234665|False",
        "4|174.140237|-59.140237|False"
      ],
      "interpretation": "The current floor pair identifies the lowest loss channel. Because
 no pair is below the loss threshold, it is diagnostic rather than the active global target
."
    },
    "weld_min_loss_limiting_pairs": {
      "lower_tube_min_pair_index": 2,
      "lower_tube_min_loss": 101.64923772970258,
      "upper_tube_min_pair_index": 2,
      "upper_tube_min_loss": 65.78401509496301,
      "same_pair_limits_both": true,
      "weld_min_loss_from_channel_mins": 167.4332528246656
    },
    "tube_pair_rankings": {
      "lower_tube_loss_weakest_to_strongest": [
        2,
        1,
        4,
        3
      ],
      "upper_tube_loss_weakest_to_strongest": [
        2,
        3,
        1,
        4
      ],
      "pair_sum_weakest_to_strongest": [
        2,
        1,
        3,
        4
```

Prompt: Verified Round Analyst *(continued)*


```
      ]
    },
    "b_proxy_summary": {
      "min_proxy_index": null,
      "max_proxy_index": null,
      "min_b_value": null,
      "max_b_value": null,
      "spread": null
    }
  },
  "coil_geometry": {
    "point_count": 11,
    "valid": true,
    "start": [
      8.633729,
      67.757545,
      14.804653
    ],
    "end": [
      4.525385,
      10.917064,
      7.610493
    ],
    "bounding_box": {
      "xmin": 4.525385,
      "xmax": 123.219238,
      "ymin": 10.917064,
      "ymax": 67.757545,
      "zmin": 1.44396,
      "zmax": 32.800928
    },
    "unique_z_values": [
      1.444,
      5.716,
      6.074,
      7.61,
      10.292,
      14.805,
      20.65,
      20.956,
      24.448,
      29.803,
      32.801
    ]
  },
  "actual_geometry_change": {
    "source": "current_coil - previous_verified_round.coil",
    "transition": {
      "source_iteration": 98,
      "transition_type": "incumbent_restore",
      "selection_branch": "incumbent_restore",
      "selected_source": "verified_incumbent",
      "fallback_used": false,
      "round_failure_mode": "none",
      "dsl_validation_status": "not_applicable",
      "dsl_validation_errors": [],
      "restored_from_verified_round": 90,
      "restore_target": "best_so_far",
      "optimizer_dsl_summary": null
    },
    "status": "measured",
    "changed_point_count": 11,
    "changed_axes": [
      "y",
      "z"
    ],
    "mean_delta_xyz": [
      0.0,
      0.198184,
      0.672273
    ],
    "max_abs_delta_xyz": [
      0.0,
      2.990854,
      1.887432
    ],
    "max_point_displacement": 3.333294,
    "top_moved_points": [
```

Prompt: Verified Round Analyst *(continued)*


```
      {
        "point_index": 1,
        "delta_xyz": [
          0.0,
          2.795855,
          1.814951
        ],
        "displacement": 3.333294
      },
      {
        "point_index": 2,
        "delta_xyz": [
          0.0,
          2.990854,
          0.935948
        ],
        "displacement": 3.133881
      },
      {
        "point_index": 8,
        "delta_xyz": [
          0.0,
          -2.963077,
          0.0
        ],
        "displacement": 2.963077
      },
      {
        "point_index": 0,
        "delta_xyz": [
          0.0,
          1.339145,
          1.887432
        ],
        "displacement": 2.31424
      },
      {
        "point_index": 9,
        "delta_xyz": [
          0.0,
          -1.849361,
          0.0
        ],
        "displacement": 1.849361
      },
      {
        "point_index": 4,
        "delta_xyz": [
          0.0,
          0.953201,
          0.918892
        ],
        "displacement": 1.323992
      },
      {
        "point_index": 3,
        "delta_xyz": [
          0.0,
          0.503577,
          0.918892
        ],
        "displacement": 1.047832
      },
      {
        "point_index": 10,
        "delta_xyz": [
          0.0,
          -0.082885,
          0.918892
        ],
        "displacement": 0.922622
      }
    ]
  },
  "incoming_transition": {
    "source_iteration": 98,
    "transition_type": "incumbent_restore",
    "selection_branch": "incumbent_restore",
```

Prompt: Verified Round Analyst *(continued)*


```
    "selected_source": "verified_incumbent",
    "fallback_used": false,
    "round_failure_mode": "none",
    "dsl_validation_status": "not_applicable",
    "dsl_validation_errors": [],
    "restored_from_verified_round": 90,
    "restore_target": "best_so_far",
    "optimizer_dsl_summary": null
  }
}

## Physical Coupling Prior
{"evidence_priority":"valid FEM metrics > channel arrays > case_memory_index >
physical_coupling_prior","core_causal_relation":"Decreasing coil-to-target distance usually
increases local magnetic coupling, induced tube loss/heating, and B-field magnitude at that
 tube
pair/proxy; increasing distance usually decreases them.","initial_adjustment_direction":"
low tube
loss or low B proxy value -> first consider small closer-coupling for relevant groups/
points.
Excessive local B/loss -> consider farther or rebalancing.","status":"initial physics
direction",
"role_use":"Verified use: interpret whether an executed move's direction matches metric
changes."}

## Scene Values
{
  "scene_values": {
    "source": "compact active_config_context; terminology brief defines meanings",
    "scene_semantics": {
      "tube_pair_count": 4,
      "tube_pair_ordering": "x_ascending",
      "b_proxy_point_count": 5,
      "b_proxy_point_ordering": "x_ascending",
      "weld_pair_region_rule": "When len(config.targets)=tube_pair_count+1, tube_pair_i is
the
semantic welding target x-region between proxy_i and proxy_{i+1}; y/z spans come from
config.welding_region.",
      "target_stack_position_rule": "upper_tube is on +z side of b_proxy_target; lower_tube
 is on -z
side; lateral station aligns with proxy/tube-pair semantics; runtime config coordinates
remain
authoritative"
    },
    "scene_geometry": {
      "b_proxy_targets": [
        {
          "proxy_index": 1,
          "xyz": [
            38.4,
            31.41525474,
            13.05
          ]
        },
        {
          "proxy_index": 2,
          "xyz": [
            52.9,
            31.41525474,
            13.05
          ]
        },
        {
          "proxy_index": 3,
          "xyz": [
            67.4,
            31.41525474,
            13.05
          ]
        },
        {
          "proxy_index": 4,
          "xyz": [
            81.9,
            31.41525474,
            13.05
          ]
        },
```

Prompt: Verified Round Analyst *(continued)*


```
        {
          "proxy_index": 5,
          "xyz": [
            96.4,
            31.41525474,
            13.05
          ]
        }
      ],
      "weld_pair_regions": [
        {
          "tube_pair_index": 1,
          "between_proxy_indices": [
            1,
            2
          ],
          "x_span": [
            38.4,
            52.9
          ],
          "x_center": 45.65,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 1 and 2"
        },
        {
          "tube_pair_index": 2,
          "between_proxy_indices": [
            2,
            3
          ],
          "x_span": [
            52.9,
            67.4
          ],
          "x_center": 60.150000000000006,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 2 and 3"
        },
        {
          "tube_pair_index": 3,
          "between_proxy_indices": [
            3,
            4
          ],
          "x_span": [
            67.4,
            81.9
          ],
          "x_center": 74.65,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 3 and 4"
        },
        {
          "tube_pair_index": 4,
```

Prompt: Verified Round Analyst *(continued)*


```
          "between_proxy_indices": [
            4,
            5
          ],
          "x_span": [
            81.9,
            96.4
          ],
          "x_center": 89.15,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 4 and 5"
        }
      ],
      "welding_region_aabb": [
        31.5,
        103.5,
        26.5,
        36.5,
        0.0,
        31.5
      ],
      "aabb_format": "xmin,xmax,ymin,ymax,zmin,zmax",
      "aabb_source": "config.welding_region"
    },
    "coordinate_semantics": {
      "x": "left-right across the welding span",
      "y": "front-back relative to welding_region",
      "z": "height; lower z is closer to the welding/proxy plane"
    }
  },
  "active_config_context": "compacted into scene_values"
}

## Minimal Current Structure Snapshot
{
  "path_semantics": {
    "meaning": "ordered control points connected as a polyline coil path",
    "body_point_count": 11,
    "body_segment_count": 10,
    "body_start_index": 0,
    "body_end_index": 10,
    "body_endpoints_are_movable": true,
    "extras_participate_as_fixed_path": true
  },
  "scene_landmarks": {
    "welding_region_summary": {
      "source": "config.welding_region",
      "x_span": [
        31.5,
        103.5
      ],
      "y_span": [
        26.5,
        36.5
      ],
      "z_span": [
        0.0,
        31.5
      ],
      "center_xyz": [
        67.5,
        31.5,
        15.75
      ]
    },
    "magnetic_proxy_targets": [
      {
        "target_id": "proxy_target_0",
        "xyz": [
          38.4,
```

Prompt: Verified Round Analyst *(continued)*


```
          31.415255,
          13.05
        ]
      },
      {
        "target_id": "proxy_target_1",
        "xyz": [
          52.9,
          31.415255,
          13.05
        ]
      },
      {
        "target_id": "proxy_target_2",
        "xyz": [
          67.4,
          31.415255,
          13.05
        ]
      },
      {
        "target_id": "proxy_target_3",
        "xyz": [
          81.9,
          31.415255,
          13.05
        ]
      },
      {
        "target_id": "proxy_target_4",
        "xyz": [
          96.4,
          31.415255,
          13.05
        ]
      }
    ],
    "welding_tube_lanes": [
      {
        "lane_id": "tube_lane_1",
        "x_range": [
          38.4,
          52.9
        ],
        "between_proxy_targets": [
          "proxy_target_0",
          "proxy_target_1"
        ],
        "x_center": 45.65
      },
      {
        "lane_id": "tube_lane_2",
        "x_range": [
          52.9,
          67.4
        ],
        "between_proxy_targets": [
          "proxy_target_1",
          "proxy_target_2"
        ],
        "x_center": 60.15
      },
      {
        "lane_id": "tube_lane_3",
        "x_range": [
          67.4,
          81.9
        ],
        "between_proxy_targets": [
          "proxy_target_2",
          "proxy_target_3"
        ],
        "x_center": 74.65
      },
      {
        "lane_id": "tube_lane_4",
        "x_range": [
          81.9,
```

Prompt: Verified Round Analyst *(continued)*


```
          96.4
        ],
        "between_proxy_targets": [
          "proxy_target_3",
          "proxy_target_4"
        ],
        "x_center": 89.15
      }
    ]
  },
  "global_shape": {
    "bbox": {
      "xmin": 4.525385,
      "xmax": 123.219238,
      "ymin": 10.917064,
      "ymax": 67.757545,
      "zmin": 1.44396,
      "zmax": 32.800928
    },
    "z_layer_summary": [
      {
        "ordinal_from_bottom": 1,
        "label": "layer_1_from_bottom",
        "display_name": "Layer 1 coil (counted from bottom to top)",
        "z_center": 7.656746,
        "point_indices": [
          0,
          1,
          2,
          3,
          4,
          10
        ]
      },
      {
        "ordinal_from_bottom": 2,
        "label": "layer_2_from_bottom",
        "display_name": "Layer 2 coil (counted from bottom to top)",
        "z_center": 25.731247,
        "point_indices": [
          5,
          6,
          7,
          8,
          9
        ]
      }
    ],
    "segment_axis_counts": {
      "x": 4,
      "y": 4,
      "z": 2
    },
    "path_length": 556.994216
  },
  "semantic_group_summary": [
    {
      "group_id": "layer_1_from_bottom",
      "group_kind": "layer",
      "point_count": 6,
      "centroid_xyz": [
        47.687303,
        33.788255,
        7.656746
      ],
      "span_xyz": [
        118.693854,
        56.840481,
        13.360694
      ],
      "relation_summary": "layer_1_from_bottom control-point layer"
    },
    {
      "group_id": "layer_2_from_bottom",
      "group_kind": "layer",
      "point_count": 5,
      "centroid_xyz": [
        51.291583,
```

Prompt: Verified Round Analyst *(continued)*


```
        34.176589,
        25.731247
      ],
      "span_xyz": [
        105.323844,
        34.869783,
        12.151312
      ],
      "relation_summary": "layer_2_from_bottom control-point layer"
    },
    {
      "group_id": "front_of_welding_region",
      "group_kind": "welding_region_direction",
      "point_count": 5,
      "centroid_xyz": [
        51.74342,
        17.549971,
        12.88867
      ],
      "span_xyz": [
        115.665475,
        11.103853,
        23.003552
      ],
      "relation_summary": "front_of_welding_region spatial side group"
    },
    {
      "group_id": "back_of_welding_region",
      "group_kind": "welding_region_direction",
      "point_count": 4,
      "centroid_xyz": [
        63.34954,
        51.300349,
        18.658687
      ],
      "span_xyz": [
        114.585509,
        28.285491,
        26.727404
      ],
      "relation_summary": "back_of_welding_region spatial side group"
    },
    {
      "group_id": "left_side_of_welding_region",
      "group_kind": "welding_region_direction",
      "point_count": 7,
      "centroid_xyz": [
        11.212615,
        35.124702,
        16.23267
      ],
      "span_xyz": [
        15.027854,
        56.840481,
        24.086458
      ],
      "relation_summary": "left_side_of_welding_region spatial side group"
    },
    {
      "group_id": "right_side_of_welding_region",
      "group_kind": "welding_region_direction",
      "point_count": 4,
      "centroid_xyz": [
        116.023356,
        31.934889,
        15.242007
      ],
      "span_xyz": [
        13.234808,
        29.542363,
        31.356968
      ],
      "relation_summary": "right_side_of_welding_region spatial side group"
    }
  ]
}
```

Evidence Card e_t


```
{
  "verified_round_id": 99,
  "source_iteration": 98,
  "evidence_quality": "high",
  "objective_value": 3.65,
  "previous_objective_value": -1.69,
  "objective_delta": 5.34,
  "improvement_status": "improved",
  "primary_blocker": "uniformity_error",
  "observed_effect": "objective improved from -1.69 to 3.65, both
weld_min_loss (167.43) and uniformity_error (4.35) meet thresholds;
geometry changed primarily in y and z with max displacement 3.33 at
point 1; tube pair 2 remains the loss floor, but its loss is above
threshold",
  "prior_consistency": "consistent with restoring the best-so-far
design and improving the worst-metric utility",
  "transition_attribution": "restore_to_best_so_far",
  "evidence_notes": "Verified round 99 restores the best-so-far
incumbent from verified round 90 after source iteration 98; DSL
validation is not applicable to this control transition; all objective
metrics satisfy their thresholds, with weld_min_loss 52.43 above its
lower threshold and uniformity_error giving a satisfaction margin of
3.65"
}
```


G.3 Cross-Round Experience Distiller

System Prompt: Case Memory Librarian

```
Maintain a structured JSON case-memory index from verified evidence
only.  Return JSON only; Markdown serialization is deterministic
post-processing.
```

Prompt: Case Memory Librarian

```
You are the Case Memory Librarian. Maintain a compact structured JSON case_memory_index
from
verified evidence cards only. The JSON index must not exceed 6000 characters; a
deterministic
renderer serializes it as Markdown after generation. Do not propose the next DSL, output
geometry
edits, or invent new physics. A single round is evidence, not a stable law. Return JSON
only.

## Terminology Brief
{"row_format":"term|meaning|goal_or_direction|how_to_use","terms":["coil_path|ordered
connected 3D
path; adjacent points form physical coil segments|geometry state|reason about a path, not
an
unordered point cloud","objective_value|maximized utility U(P)=-F(P) computed from verified
 FEM metrics|higher_is_better|compare verified rounds; larger value is better","
weld_min_loss|min(lower_tube_loss_values) +
min(upper_tube_loss_values)|higher_is_better|track min loss","uniformity_error|average
range-based spread of lower- and upper-tube loss channels|lower_is_better|track inter-pair
loss balance",
"lower_tube_loss_values|4 lower-tube losses by tube_pair_index|feeds
lower_tube_loss_uniformity|name weak lower region","upper_tube_loss_values|4 upper-tube
losses by
tube_pair_index|feeds upper_tube_loss_uniformity|name weak upper region","verified_round|
one coil state
that has completed real FEM simulation|evidence unit|compare objective/metrics only after
real
simulation","evidence_card|compact JSON summary of one verified_round|single-round evidence
|feed
case memory and optimizer","case_memory_index|compact cross-round memory from verified
evidence
cards|case-level memory|provide SAP with experience without writing executable geometry",
```

Prompt: Case Memory Librarian *(continued)*

```
"search_policy|planning posture: exploit, explore, or escape_plateau|non-geometric metadata
|record search intent without selecting coefficients or changing physics","scale|coarse
action range resolved by GGT: micro/small/medium/large|action range|GGT maps it to bounded
coefficient intervals under feasibility limits"]}

## Metric Snapshot
{"row_format":"metric|current_value|threshold|direction|display_name|margin_to_threshold",
"metrics":["weld_min_loss||115.0|higher_is_better|Welding Loss|",
"uniformity_error||8.0|lower_is_better|Loss Uniformity Error|"]}

## Physical Coupling Prior
{"evidence_priority":"valid FEM metrics > channel arrays > case_memory_index >
physical_coupling_prior","core_causal_relation":"Decreasing coil-to-target distance usually
increases local magnetic coupling, induced tube loss/heating, and B-field magnitude at that
 tube
pair/proxy; increasing distance usually decreases them.","initial_adjustment_direction":"
low tube
loss or low B proxy value -> first consider small closer-coupling for relevant groups/
points.
Excessive local B/loss -> consider farther or rebalancing.","status":"initial physics
direction",
"role_use":"Librarian use: compact repeated verified evidence into a case memory index; do
not
invent new physical laws."}

## Output JSON Schema
{
  "best_so_far": {
    "verified_round_id": "int|null",
    "objective_value": "number|null",
    "metric_profile": {}
  },
  "repeated_verified_observations": [
    "max 5 evidence-backed facts"
  ],
  "failed_or_noop_patterns": [
    "max 5 repeated failed/no-op patterns"
  ],
  "search_bias_for_optimizer": {
    "preferred_policy": "exploit | explore | escape_plateau",
    "preferred_scale": "micro | small | medium | large",
    "preferred_groups": [],
    "avoid_patterns": []
  },
  "recent_evidence_digest": [
    "max 5 compact verified evidence notes"
  ]
}

## Scene Values
{
  "scene_values": {
    "source": "compact active_config_context; terminology brief defines meanings",
    "scene_semantics": {
      "tube_pair_count": 4,
      "tube_pair_ordering": "x_ascending",
      "b_proxy_point_count": 5,
      "b_proxy_point_ordering": "x_ascending",
      "weld_pair_region_rule": "When len(config.targets)=tube_pair_count+1, tube_pair_i is
the
semantic welding target x-region between proxy_i and proxy_{i+1}; y/z spans come from
config.welding_region.",
      "target_stack_position_rule": "upper_tube is on +z side of b_proxy_target; lower_tube
 is on -z
side; lateral station aligns with proxy/tube-pair semantics; runtime config coordinates
remain
authoritative"
    },
    "scene_geometry": {
      "b_proxy_targets": [
        {
          "proxy_index": 1,
          "xyz": [
            38.4,
            31.41525474,
            13.05
          ]
        },
```

Prompt: Case Memory Librarian *(continued)*


```
        {
          "proxy_index": 2,
          "xyz": [
            52.9,
            31.41525474,
            13.05
          ]
        },
        {
          "proxy_index": 3,
          "xyz": [
            67.4,
            31.41525474,
            13.05
          ]
        },
        {
          "proxy_index": 4,
          "xyz": [
            81.9,
            31.41525474,
            13.05
          ]
        },
        {
          "proxy_index": 5,
          "xyz": [
            96.4,
            31.41525474,
            13.05
          ]
        }
      ],
      "weld_pair_regions": [
        {
          "tube_pair_index": 1,
          "between_proxy_indices": [
            1,
            2
          ],
          "x_span": [
            38.4,
            52.9
          ],
          "x_center": 45.65,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 1 and 2"
        },
        {
          "tube_pair_index": 2,
          "between_proxy_indices": [
            2,
            3
          ],
          "x_span": [
            52.9,
            67.4
          ],
          "x_center": 60.150000000000006,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 2 and 3"
        },
```

Prompt: Case Memory Librarian *(continued)*

```
        {
          "tube_pair_index": 3,
          "between_proxy_indices": [
            3,
            4
          ],
          "x_span": [
            67.4,
            81.9
          ],
          "x_center": 74.65,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 3 and 4"
        },
        {
          "tube_pair_index": 4,
          "between_proxy_indices": [
            4,
            5
          ],
          "x_span": [
            81.9,
            96.4
          ],
          "x_center": 89.15,
          "y_span": [
            26.5,
            36.5
          ],
          "z_span": [
            0.0,
            31.5
          ],
          "source": "derived_from_adjacent_config_targets",
          "meaning": "semantic welding target region between B proxy 4 and 5"
        }
      ],
      "welding_region_aabb": [
        31.5,
        103.5,
        26.5,
        36.5,
        0.0,
        31.5
      ],
      "aabb_format": "xmin,xmax,ymin,ymax,zmin,zmax",
      "aabb_source": "config.welding_region"
    },
    "coordinate_semantics": {
      "x": "left-right across the welding span",
      "y": "front-back relative to welding_region",
      "z": "height; lower z is closer to the welding/proxy plane"
    }
  },
  "active_config_context": "compacted into scene_values"
}

## Global knowledge
{
  "markdown": "# Scenario and Priors\n\n## Scene Semantics\n- The coil is an ordered
connected 3D
path, not an unordered point cloud.\n- Adjacent coordinates define consecutive physical
coil
segments.\n- The welding assembly, welding region, and B proxy targets are loaded from
runtime
config.\n- Tube-pair regions are semantic welding target x-regions derived from adjacent B
proxy
targets when `len(config.targets) = tube_pair_count + 1`.\n- Runtime config is
authoritative for
```

Prompt: Case Memory Librarian *(continued)*

```
coordinates. Static prose must not override ‘config.targets‘ or ‘config.welding_region‘.\n\
n##
Objective Semantics\n- Verified FEM metrics are the source of truth. ‘objective_value‘
denotes the maximized utility ‘U(P)=-F(P)‘, so higher is better.\n- Metric directions and
thresholds come from ‘config.metrics.simulation_evaluation‘.\n-
B proxy metrics are diagnostic evidence; they do not outrank the verified utility.\n- A
single verified round is evidence, not a stable law.\n- Repeated verified evidence is
required
before storing a reusable case-level pattern.\n\n## Physical Coupling Prior\n- Moving coil
regions
closer to the welding region usually increases local magnetic coupling, induced tube loss,
and
local field strength.\n[truncated]"
}

## Current Case Memory Index
{
  "markdown": "# Case Memory Index\n\n## Best So Far\n- verified_round_id: 90\n-
objective_value: 3.65\n-
metric_profile: {’weld_min_loss’: 167.43, ’uniformity_error’: 4.35}\n\n## Repeated Verified
Observations\n- Small-scale y moves toward welding region from non-positive objective (<=0)
 yield consistent
gains (rounds 93,94).\n- Small-scale y moves toward welding region from best objective
(>=3.65) cause objective drop (round 91).\n- Small-scale y moves toward welding region from
moderate state (1.0-2.5) can cause drop if uniformity is near threshold (rounds 96,97).\n-
Tube
pairs 2 and 3 most frequently have the lowest weld-loss responses (rounds 89-98).\n- Z-axis
downward moves (negative z) from low state improved objective (round 98) but uniformity
remained
problematic.\n\n## Failed Or No-op Patterns\n- Small-scale y move toward welding region
from best
objective (round 91).\n- Simultaneous y moves of layers 1&2 caused severe drop (round 92).\
n-
Small-scale y move toward welding region from moderate state with uniformity near threshold
 (rounds
96,97).\n- Medium-scale y moves from moderate-good state (round 88).\n- Z-axis upward moves
 (any
scale) remain failed historically.\n\n## Search Bias For Optimizer\n- preferred_policy:
explore\n-
preferred_scale: small\n- preferred_groups: []\n- avoid_patterns: [’small y move from best
state’,
’small y move from moderate state if uniformity near threshold’, ’simultaneous y move of
layers
1&2’, ’medium y move from moderate-good state’, ’z upward move’]\n\n## Recent Evidence
Digest\n-
Round 89: Small y front_of_welding_region from -1.58 to +4.95 delta.\n- Round 90: Small y
back_of_welding_region from 3.37 to 3.65 (best).\n- Round 91: Small y
back_of_welding_region from
3.65 dropped to 2.65.\n- Round 92: Simultaneous layers 1&2 y moves from 2.65 to -2.38.\n[
truncated]"
}

## Recent Verified Evidence Cards
{
  "markdown": "# Round Ledger\n\nCase: ‘case_20260703_022328_deepseek-v4-flash‘\n\nThis
file is
recreated for each case run.\nIt records round-by-round simulation analysis, semantic-
action DSL, BO
selection outcomes, and round-level experience for this experiment only.\n\n### Round 90\n-
Objective snapshot: objective=3.65; parse_status=ok\n- Previous objective: 3.37\n-
Objective delta:
0.28\n- Improvement status: improved\n- Primary blocker: uniformity_error\n- Strategy
applied: The
current geometry is attributed to the DSL transition from iteration 89, selection branch
’sge_best_query’, selected source ’sge_query’, with no fallback and DSL validation
status ’accepted’. The DSL proposal was a semantic...\n- Verification context:
source_iteration=89;
selection_branch=sge_best_query; selected_source=sge_query; fallback_used=false;
round_failure_mode=none; simulation_cache_hit=true; simulation_cache_source=sge_query;
restored_from_iteration=u...\n- DSL/SGE outcome: source_iteration=89; dsl_validation=
accepted;
sge_best_objective=3.65; sge_simulation_count=1; objective_delta=0.28; improvement_status=
improved\n-
Channel metric analysis: Tube pair 2 has the minimum lower- and upper-tube loss values;
lower_tube_min_pair_index=2;
upper_tube_min_pair_index=2; same_pair_limits_both=true; b_min_proxy_index=unavailable;...\
n-
```

Prompt: Case Memory Librarian *(continued)*

```
Actionable insight: Objective improved by 0.28 (from 3.37 to 3.65). Both weld_min_loss
(167.43,
threshold 115, meets threshold) and uniformity_error (4.35, threshold 8, meets threshold)
are
within acceptable bounds. The geometry change moved 4 points in the back_of_welding_region
primarily along the y-axis (mean delta y = -0.10, max displacement 0.89). Point 1 moved
-0.90 y
(to...\n\n### Round 91\n- Objective snapshot: objective=2.65; parse_status=ok\n- Previous
objective: 3.65\n- Objective delta: -1\n- Improvement status: worsened\n- Primary blocker:
uniformity_error\n- Strategy applied: sge_query_accepted_worse with semantic_trust_region
DSL: back_of_welding_region y-axis toward welding region, exploit policy.\n- Verification
context:
source_iteration=90; selection_branch=sge_query_accepted_worse;
selected_source=sge_query; fallback_used=false; round_failure_mode=none;
simulation_cache_hit=true; simulation_cache_source=sge_query; resto...\n- DSL/SGE
outcome: source_iteration=90; dsl_validation=accepted; sge_best_objective=2.65;
sge_simulation_count=1; objective_delta=-1; improvement_status=worsened\n- Channel metric
analysis:
Lower tube pair 3 and upper tube pair 2 currently limit weld_min_loss; target geometry
should
improve these weak channels without worsening channel balance.; lower_tube_min_pair_index
=3;
upper_tube_min_pair_index=2; same_pair_limits_both=fals...\n- Actionable insight: Objective
 worsened
by 1.0; weld_min_loss (170.81) and uniformity_error (5.35) still meet their thresholds;
limiting pairs
are lower tube pair 3 and upper tube pair 2.\n\n### Round 92\n- Objective snapshot:
objective=-2.38;
parse_status=ok\n- Previous objective: 2.65\n- Objective delta: -5.03\n- Improvement status
:
worsened\n- Primary blocker: uniformity_error\n- Strategy applied: Attributed to
sge_query_accepted_worse from iteration 91. DSL: semantic_trust_region on layers 1&2,
y-axis, toward welding region.\n- Verification context: source_iteration=91;
selection_branch=sge_query_accepted_worse; selected_source=sge_query;
fallback_used=false; round_failure_mode=none; simulation_cache_hit=true;
simulation_cache_source=sge_query; resto...\n- DSL/SGE outcome: source_iteration=91;
dsl_validation=accepted; sge_best_objective=-2.38; sge_simulation_count=1; objective_delta
=-5.03;
improvement_status=worsened\n- Channel metric analysis: Lower tube pair 2 and upper tube
pair 3
currently limit weld_min_loss; target geometry should improve these weak channels without
worsening
channel balance.; lower_tube_min_pair_index=2; upper_tube_min_pair_index=3;
same_pair_limits_both=fals...\n- Actionable insight: Moving points in layer_1 negatively
and
layer_2 positively in y (toward and away from welding region) resulted in worsened
uniformity error
(10.38 vs 8.0 threshold) while weld_min_loss remained above threshold. Overall objective
dropped.\n\n### Round 93\n- Objective snapshot: objective=-0.07; parse_status=ok\n-
Previous
objective: -2.38\n- Objective delta: 2.31\n- Improvement status: improved\n- Primary
blocker:
uniformity_error\n- Strategy applied: Geometry is the result of the SGE
query ranked best from iteration 92 (sge_best_query). The DSL proposed a small-scale
exploit move on front_of_welding_region in the y-axis toward welding region, target...\n-
Verification context: source_iteration=92; selection_branch=sge_best_query;
selected_source=sge_query; fallback_used=false; round_failure_mode=none;
simulation_cache_hit=true; simulation_cache_source=sge_query;
restored_from_iteration=u...\n- DSL/SGE outcome: source_iteration=92; dsl_validation=
accepted;
sge_best_objective=-0.07; sge_simulation_count=1; objective_delta=2.31;
improvement_status=improved\n- Channel metric analysis: Lower tube pair 2 and upper tube
pair 3
currently limit weld_min_loss; target geometry should improve these weak channels without
worsening
channel balance.; lower_tube_min_pair_index=2; upper_tube_min_pair_index=3;
same_pair_limits_both=fals...\n- Actionable insight: Moved 5 points in
front_of_welding_region
group +y (toward welding region) with max displacement 1.94. Objective improved
significantly
(delta 2.31); weld_min_loss (166.07) well above threshold (115.0); uniformity_error
(8.06813)
slightly violates threshold (8.0) by 0.06813.\n\n### Round 94\n- Objective snapshot:
objective=0.98; parse_status=ok\n- Previous objective: -0.07\n- Objective delta: 1.05\n-
Improvement status: improved\n- Primary blocker: uniformity_error\n- Strategy applied:
Attributed to
sge_best_query selected DSL with proposal type semantic_trust_region, exploit policy,
focusing on
```

Prompt: Case Memory Librarian *(continued)*

```
front_of_welding_region y-axis with small moves. DSL validation accepted, no fallback.\n-
Verification context: source_iteration=93; selection_branch=sge_best_query;
selected_source=sge_query; fallback_used=false; round_failure_mode=none;
simulation_cache_hit=true; simulation_cache_source=sge_query;
restored_from_iteration=u...\n- DSL/SGE outcome: source_iteration=93; dsl_validation=
accepted;
sge_best_objective=0.98; sge_simulation_count=1; objective_delta=1.05; improvement_status=
improved\n-
Channel metric analysis: Lower tube pair 2 and upper tube pair 3 currently limit
weld_min_loss;
target geometry should improve these weak channels without worsening channel balance.;
lower_tube_min_pair_index=2; upper_tube_min_pair_index=3; same_pair_limits_both=fals...\n-
Actionable insight: Small y-axis adjustments (max 0.77) in front_of_welding_region improved
objective dramatically from -0.07 to 0.98. Both weld_min_loss (163.95 > 115) and
uniformity_error
(7.02 < 8) now meet thresholds. Weld_min_loss increased sufficiently above threshold;
uniformity
improved. Weakest channels: lower tube pair 2 and upper tube pair 3.\n\n### Round 95\n-
Objective
snapshot: objective=1.24; parse_status=ok\n- Previous objective: 0.98\n- Objective delta:
0.26\n-
Improvement status: improved\n- Primary blocker: uniformity_error\n- Strategy applied:
Attributed to
SGE query (iteration 94) selected via sge_best_query branch, no fallback, dsl
accepted.\n- Verification context: source_iteration=94; selection_branch=sge_best_query;
selected_source=sge_query; fallback_used=false; round_failure_mode=none;
simulation_cache_hit=true; simulation_cache_source=sge_query;
restored_from_iteration=u...\n- DSL/SGE outcome: source_iteration=94; dsl_validation=
accepted;
sge_best_objective=1.24; sge_simulation_count=1; objective_delta=0.26; improvement_status=
improved\n-
Channel metric analysis: Tube pair 2 has the minimum lower- and upper-tube loss values;
lower_tube_min_pair_index=2;
upper_tube_min_pair_index=2; same_pair_limits_both=true; b_min_proxy_index=unavailable;...\
n-
Actionable insight: Objective improved by +0.26 to 1.24, driven by a semantic trust region
move
along the y-axis (front/back relative to welding region) that primarily adjusted points in
front
and back groups. Both weld_min_loss (156.8, meets threshold 115) and uniformity_error
(6.76, meets
threshold 8) improved relative to previous round. Tube pair 2 remains the limiting floor
...\n\n###
Round 96\n- Objective snapshot: objective=1.06; parse_status=ok\n- Previous objective:
1.24\n-
Objective delta: -0.18\n- Improvement status: worsened\n- Primary blocker: uniformity_error
\n-
Strategy applied: Selected SGE query (semantic_trust_region, exploit, small y moves
toward welding region for front/back groups) after acceptance as worse result.\n-
Verification
context: source_iteration=95; selection_branch=sge_query_accepted_worse;
selected_source=sge_query; fallback_used=false; round_failure_mode=none;
simulation_cache_hit=true; simulation_cache_source=sge_query; resto...\n- DSL/SGE
outcome: source_iteration=95; dsl_validation=accepted; sge_best_objective=1.06;
sge_simulation_count=1; objective_delta=-0.18; improvement_status=worsened\n- Channel
metric
analysis: Lower tube pair 2 and upper tube pair 3 currently limit weld_min_loss; target
geometry
should improve these weak channels without worsening channel balance.;
lower_tube_min_pair_index=2;
upper_tube_min_pair_index=3; same_pair_limits_both=fals...\n- Actionable insight: Objective
decreased despite moving front and back coil points toward welding region; both metrics
still met
their thresholds, but the worst-metric margin decreased.\n\n### Round 97\n- Objective
snapshot: objective=-2.45; parse_status=ok\n- Previous objective: 1.06\n- Objective delta:
-3.51\n-
Improvement status: worsened\n- Primary blocker: uniformity_error\n- Strategy applied: The
geometry
was updated by the optimizer's sge_query (Proposal: semantic_trust_region focusing on
weld_min_loss in front_of_welding_region, y-axis exploit, small scale). The DSL was
validated and
executed without failure or fal...\n- Verification context: source_iteration=96;
selection_branch=sge_query_accepted_worse; selected_source=sge_query;
fallback_used=false; round_failure_mode=none; simulation_cache_hit=true;
simulation_cache_source=sge_query; resto...\n- DSL/SGE outcome: source_iteration=96;
dsl_validation=accepted; sge_best_objective=-2.45; sge_simulation_count=1; objective_delta
=-3.51;
```

Prompt: Case Memory Librarian *(continued)*

```
improvement_status=worsened\n- Channel metric analysis: Tube pair 2 has the minimum lower-
and upper-tube loss values;
lower_tube_min_pair_index=2; upper_tube_min_pair_index=2; same_pair_limits_both=true;
b_min_proxy_index=unavailable;...\n- Actionable insight: The SGE query, which exploited
y-axis adjustments in the front_of_welding_region to improve weld_min_loss, was accepted
despite
worsening the overall objective. While weld_min_loss (154.31) exceeds its threshold of 115,
uniformity_error (10.45) violated its threshold of 8, increasing from the prior round. Tube
 pair 2
remains the limiting pair for...\n\n### Round 98\n- Objective snapshot: objective=-1.69;
parse_status=ok\n- Previous objective: -2.45\n- Objective delta: 0.76\n- Improvement status
:
improved\n- Primary blocker: uniformity_error\n- Strategy applied: Geometry from iteration
98
originated from sge_query selection of iteration 97's semantic trust region update
targeting uniformity_error via exploit policy on layer_1_from_bottom in negative z
direction.\n-
Verification context: source_iteration=97; selection_branch=sge_best_query;
selected_source=sge_query; fallback_used=false; round_failure_mode=none;
simulation_cache_hit=true; simulation_cache_source=sge_query;
restored_from_iteration=u...\n- DSL/SGE outcome: source_iteration=97; dsl_validation=
accepted;
sge_best_objective=-1.69; sge_simulation_count=1; objective_delta=0.76;
improvement_status=improved\n- Channel metric analysis: Lower tube pair 2 and upper tube
pair 1
currently limit weld_min_loss; target geometry should improve these weak channels without
worsening
channel balance.; lower_tube_min_pair_index=2; upper_tube_min_pair_index=1;
same_pair_limits_both=fals...\n- Actionable insight: Lowering layer_1_from_bottom points (
mean
delta z = -0.672) improved objective by 0.76, but uniformity_error remains the primary
blocker
(margin -1.685). Weak tube pairs (lower pair 2, upper pair 1) persist as the limiting
channels.\n\n### Round 99\n- Objective snapshot: objective=3.65; parse_status=ok\n-
Previous
objective: -1.69\n- Objective delta: 5.34\n- Improvement status: improved\n- Primary
blocker: uniformity_error\n- Strategy applied: restore_to_best_so_far\n- Verification
context:
source_iteration=98; transition_type=incumbent_restore; selection_branch=incumbent_restore;
 selected_source=verified_incumbent;
fallback_used=false; round_failure_mode=none; simulation_cache_hit=true;
simulation_cache_source=verified_incumbent; restored_from_verified_round=90\n- DSL/SGE
outcome:
source_iteration=98; dsl_validation=not_applicable; sge_best_objective=not_applicable;
sge_simulation_count=not_applicable; objective_delta=5.34; improvement_status=improved\n-
Channel
metric analysis: Tube pair 2 has the minimum lower- and upper-tube loss values;
lower_tube_min_pair_index=2;
upper_tube_min_pair_index=2; same_pair_limits_both=true; b_min_proxy_index=unavailable;...\
n-
Actionable insight: objective improved from -1.69 to 3.65, both weld_min_loss (167.43) and
uniformity_error (4.35) meet thresholds; geometry changed primarily in y and z with max
displacement 3.33 at point 1; tube pair 2 remains the loss floor, but its loss is above
threshold"
}
```

Distilled Case Memory $\mathcal{E}_t$

```
{
  "best_so_far": {
    "verified_round_id": 90,
    "objective_value": 3.65,
    "metric_profile": {
      "weld_min_loss": 167.43,
      "uniformity_error": 4.35
    }
  },
  "repeated_verified_observations": [
    "Small-scale y moves toward welding region from non-positive
objective (<=0) are associated with consistent gains (rounds 93,94).",
    "Small-scale y moves toward welding region from best objective
(3.65) are associated with an objective drop (round 91).",
```

Distilled Case Memory $\mathcal{E}_t$ *(continued)*


```
    "Small-scale y moves toward welding region from moderate state
(1.0-2.5) are associated with drops when uniformity is near threshold (
rounds 96,97).",
    "Tube pair 2 is the most frequent weakest-loss pair (rounds
90,95,97,99).",
    "Z-axis downward moves (negative z) from low state coincided with
an objective improvement (round 98) but uniformity remained problematic
."
  ],
  "failed_or_noop_patterns": [
    "Small y move toward welding region from best objective (round 91)
.",
    "Simultaneous y moves of layers 1&2 coincided with a severe drop (
round 92).",
    "Small y move toward welding region from moderate state with
uniformity near threshold (rounds 96,97).",
    "Medium-scale y moves from moderate-good state (round 88).",
    "Observed z-axis upward moves remain unsuccessful in the recorded
history."
  ],
  "search_bias_for_optimizer": {
    "preferred_policy": "explore",
    "preferred_scale": "small",
    "preferred_groups": [],
    "avoid_patterns": [
      "small y move from best state",
      "small y move from moderate state if uniformity near threshold",
      "simultaneous y move of layers 1&2",
      "medium y move from moderate-good state",
      "z upward move"
    ]
  },
  "recent_evidence_digest": [
    "Round 90: small y back_of_welding_region from 3.37 to 3.65 (best)
.",
    "Round 91: small y back_of_welding_region from 3.65 dropped to
2.65.",
    "Round 92: simultaneous layers 1&2 y moves from 2.65 to -2.38.",
    "Round 93: small y front_of_welding_region from -2.38 to -0.07.",
    "Round 99: restore_to_best_so_far achieved 3.65 again."
  ]
}
```


G.4 Semantic Action Planner

System Prompt: Semantic Action Planner

```
You optimize 3D coils.  Return exactly one valid semantic-action DSL
object.
```

Prompt: Semantic Action Planner

```
You are the Semantic Action Planner. Return ONE compact JSON semantic-action DSL only.
Specify the deformation operation, focus, direction, physical rationale, and coarse range.
GGT constructs the executable deformation family, and SGE determines its coefficients. Do
not output point coordinates, point indices, geometry validation instructions, or multiple
alternatives.

## Terminology Brief
{"row_format":"term|meaning|goal_or_direction|how_to_use","terms":["coil_path|ordered
connected 3D
path; adjacent points form physical coil segments|geometry state|reason about a path, not
an
```

Prompt: Semantic Action Planner *(continued)*

```
unordered point cloud","objective_value|maximized utility U(P)=-F(P) computed from verified
 FEM metrics|higher_is_better|compare verified rounds; larger value is better","
weld_min_loss|min(lower_tube_loss_values) +
min(upper_tube_loss_values)|higher_is_better|track min loss","uniformity_error|average
range-based spread of lower- and upper-tube loss channels|lower_is_better|track inter-pair
loss balance",
"lower_tube_loss_values|4 lower-tube losses by tube_pair_index|feeds
lower_tube_loss_uniformity|name weak lower region","upper_tube_loss_values|4 upper-tube
losses by
tube_pair_index|feeds upper_tube_loss_uniformity|name weak upper region","weld_pair_region|
semantic
welding target x-region between adjacent B proxy targets|true welding target region|map
tube-pair
loss channels to spatial x spans between proxy points","welding_region_aabb|welding region
AABB;
spatial reference, not an optimization objective; welding_region_aabb z span is [0.0,
31.5];
config.welding_region=[31.5, 103.5, 26.5, 36.5, 0.0, 31.5]|spatial reference|classify
sectors and
interpret toward/away-from-welding-region movement","verified_round|one coil state that has
completed real FEM simulation|evidence unit|compare objective/metrics only after real
simulation",
"evidence_card|compact JSON summary of one verified_round|single-round evidence|feed case
memory
and optimizer","case_memory_index|compact cross-round memory from verified evidence
cards|case-level memory|provide SAP with experience without writing executable geometry",
"semantic_trust_region|DSL tag for a coordinated translation request|GGT input|select
semantic groups, axes, direction, and coarse range; GGT constructs the bounded family","
semantic_group|compiler-defined subset
of body points such as layer/front/back/left/right|action focus|select where SGE may search
without naming raw point indices","search_policy|planning posture: exploit, explore, or
escape_plateau|non-geometric metadata|record search intent without selecting coefficients
or changing physics","scale|coarse action range resolved by GGT: micro/small/medium/large|
action range|GGT maps it to bounded coefficient intervals under feasibility limits",
"direction|symbolic preferred direction resolved by translator|mode orientation|use
positive/negative/either/toward_welding_region/away_from_welding_region; the symmetric
coefficient box permits local refinement in either sign"]}

## Metric Snapshot
{"row_format":"metric|current_value|threshold|direction|display_name|margin_to_threshold",
"metrics":["weld_min_loss|167.4332528246656|115.0|higher_is_better|Welding Loss
|52.4332528246656","uniformity_error|4.351902442438149|8.0|lower_is_better|Loss Uniformity
Error|3.648097557561851"]}

## Decision Brief
{"primary_blocker":"uniformity_error","objective_value":3.65,"improvement_status":"improved
",
"best_so_far":{"best_verified_round":90,"best_objective_value":3.65,"current_verified_round
":99,
"current_objective_value":3.65,"objective_drop_from_best":0.0,"no_improve_rounds":0,
"restore_verified_round":90,"best_metric_snapshot":"objective_value=3.65|weld_min_loss
=167.4332528246656|
uniformity_error=4.351902442438149"},"channel_focus":"Uniformity error is the active worst-
metric term. Tube pair 2 has the minimum lower- and upper-tube losses, but all pairwise
loss responses remain above the loss threshold.",
"available_semantic_groups":["layer_1_from_bottom","layer_2_from_bottom","
front_of_welding_region",
"back_of_welding_region","left_side_of_welding_region","right_side_of_welding_region"]}

## Causal Priors
{"evidence_priority":"verified simulation metrics > SGE evidence > channel evidence >
physical prior > librarian prose","coupling_loss_prior":"Closer coil-to-welding-region
coupling
usually increases local tube loss/heating; farther usually decreases it.",
"collective_loss_prior":"If multiple tube pairs are under target, prefer collective group-
level
closer-coupling, not isolated one-by-one moves.","proxy_plane_prior":"Farther from the
proxy plane
usually reduces B sensitivity/non-uniformity; closer may increase B magnitude/spread."}

## Scene Values
{"tube_pair_count":4,"tube_pair_ordering":"x_ascending","b_proxy_point_count":5,
"b_proxy_point_ordering":"x_ascending","target_stack_position_rule":"upper_tube is on +z
side of
b_proxy_target; lower_tube is on -z side; lateral station aligns with proxy/tube-pair
semantics;
runtime config coordinates remain authoritative","b_proxy_targets":[{"proxy_index":1,"xyz
":[38.4,
```

Prompt: Semantic Action Planner *(continued)*


```
31.41525474,13.05]},{"proxy_index":2,"xyz":[52.9,31.41525474,13.05]},{"proxy_index":3,"xyz
":[67.4,
31.41525474,13.05]},{"proxy_index":4,"xyz":[81.9,31.41525474,13.05]},{"proxy_index":5,"xyz
":[96.4,
31.41525474,13.05]}],"weld_pair_regions":[{"tube_pair_index":1,"between_proxy_indices
":[1,2],
"x_span":[38.4,52.9],"x_center":45.65,"y_span":[26.5,36.5],"z_span":[0.0,31.5],
"source":"derived_from_adjacent_config_targets","meaning":"semantic welding target region
between B
proxy 1 and 2"},{"tube_pair_index":2,"between_proxy_indices":[2,3],"x_span":[52.9,67.4],
"x_center":60.150000000000006,"y_span":[26.5,36.5],"z_span":[0.0,31.5],
"source":"derived_from_adjacent_config_targets","meaning":"semantic welding target region
between B
proxy 2 and 3"},{"tube_pair_index":3,"between_proxy_indices":[3,4],"x_span":[67.4,81.9],
"x_center":74.65,"y_span":[26.5,36.5],"z_span":[0.0,31.5],
"source":"derived_from_adjacent_config_targets","meaning":"semantic welding target region
between B
proxy 3 and 4"},{"tube_pair_index":4,"between_proxy_indices":[4,5],"x_span":[81.9,96.4],
"x_center":89.15,"y_span":[26.5,36.5],"z_span":[0.0,31.5],
"source":"derived_from_adjacent_config_targets","meaning":"semantic welding target region
between B
proxy 4 and 5"}],"coordinate_semantics":{"x":"left-right across welding span","y":"front-
back
relative to welding region","z":"height; lower usually closer to welding/proxy plane"}}

## Current Coil Structure
{"path_semantics":{"meaning":"ordered control points connected as a polyline coil path",
"body_point_count":11,"body_segment_count":10,"body_start_index":0,"body_end_index":10,
"body_endpoints_are_movable":true,"extras_participate_as_fixed_path":true},
"scene_landmarks":{"welding_region_summary":{"source":"config.welding_region","x_span
":[31.5,103.5],
"y_span":[26.5,36.5],"z_span":[0.0,31.5],"center_xyz":[67.5,31.5,15.75]},
"magnetic_proxy_targets":[{"target_id":"proxy_target_0","xyz":[38.4,31.415255,13.05]},
{"target_id":"proxy_target_1","xyz":[52.9,31.415255,13.05]},{"target_id":"proxy_target_2",
"xyz":[67.4,31.415255,13.05]},{"target_id":"proxy_target_3","xyz":[81.9,31.415255,13.05]},
{"target_id":"proxy_target_4","xyz":[96.4,31.415255,13.05]}],
"welding_tube_lanes":[{"lane_id":"tube_lane_1","x_range":[38.4,52.9],
"between_proxy_targets":["proxy_target_0","proxy_target_1"],"x_center":45.65},
{"lane_id":"tube_lane_2","x_range":[52.9,67.4],"between_proxy_targets":["proxy_target_1",
"proxy_target_2"],"x_center":60.15},{"lane_id":"tube_lane_3","x_range":[67.4,81.9],
"between_proxy_targets":["proxy_target_2","proxy_target_3"],"x_center":74.65},
{"lane_id":"tube_lane_4","x_range":[81.9,96.4],"between_proxy_targets":["proxy_target_3",
"proxy_target_4"],"x_center":89.15}]},"global_shape":{"bbox":{"xmin":4.525385,"xmax
":123.219238,
"ymin":10.917064,"ymax":67.757545,"zmin":1.44396,"zmax":32.800928},
"z_layer_summary":[{"ordinal_from_bottom":1,"label":"layer_1_from_bottom","display_name":"
Layer 1
coil (counted from bottom to top)","z_center":7.656746,"point_indices":[0,1,2,3,4,10]},
{"ordinal_from_bottom":2,"label":"layer_2_from_bottom","display_name":"Layer 2 coil (
counted from
bottom to top)","z_center":25.731247,"point_indices":[5,6,7,8,9]}],"segment_axis_counts":{"
x":4,
"y":4,"z":2},"path_length":556.994216},"semantic_groups":[{"group_id":"layer_1_from_bottom
",
"group_kind":"layer","point_count":6,"centroid_xyz":[47.687303,33.788255,7.656746],
"span_xyz":[118.693854,56.840481,13.360694],"relation_summary":"layer_1_from_bottom control
-point
layer"},{"group_id":"layer_2_from_bottom","group_kind":"layer","point_count":5,
"centroid_xyz":[51.291583,34.176589,25.731247],"span_xyz":[105.323844,34.869783,12.151312],
"relation_summary":"layer_2_from_bottom control-point layer"},{"group_id":"
front_of_welding_region",
"group_kind":"welding_region_direction","point_count":5,"centroid_xyz
":[51.74342,17.549971,12.88867]
,"span_xyz":[115.665475,11.103853,23.003552],"relation_summary":"front_of_welding_region
spatial
side group"},{"group_id":"back_of_welding_region","group_kind":"welding_region_direction",
"point_count":4,"centroid_xyz":[63.34954,51.300349,18.658687],"span_xyz
":[114.585509,28.285491,
26.727404],"relation_summary":"back_of_welding_region spatial side group"},
{"group_id":"left_side_of_welding_region","group_kind":"welding_region_direction","
point_count":7,
"centroid_xyz":[11.212615,35.124702,16.23267],"span_xyz":[15.027854,56.840481,24.086458],
"relation_summary":"left_side_of_welding_region spatial side group"},
{"group_id":"right_side_of_welding_region","group_kind":"welding_region_direction","
point_count":4,
"centroid_xyz":[116.023356,31.934889,15.242007],"span_xyz":[13.234808,29.542363,31.356968],
"relation_summary":"right_side_of_welding_region spatial side group"}],
"point_semantics_compact":["0|[8.633729, 67.757545, 14.804653]
```

Prompt: Semantic Action Planner *(continued)*

```
|layer_1_from_bottom|back_of_welding_region|left_side_of_welding_region|above_proxy_plane|
tube_lane_
1|proxy_target_0|endpoint|layer_1_from_bottom,back_of_welding_region,
left_side_of_welding_region",
"1|[123.219238, 39.472054, 6.073523]|layer_1_from_bottom|back_of_welding_region|
right_side_of_weldin
g_region|below_proxy_plane|tube_lane_4|proxy_target_4|corner|layer_1_from_bottom,
back_of_welding_region,right_side_of_welding_region","2|[120.190859, 21.252224, 1.44396]
|layer_1_from_bottom|front_of_welding_region|right_side_of_welding_region|below_proxy_plane
|tube_lan
e_4|proxy_target_4|corner|layer_1_from_bottom,front_of_welding_region,
right_side_of_welding_region",
"3|[18.641372, 22.020917, 10.29177]|layer_1_from_bottom|front_of_welding_region|
left_side_of_welding
_region|below_proxy_plane|tube_lane_1|proxy_target_0|corner|layer_1_from_bottom,
front_of_welding_region,left_side_of_welding_region","4|[10.913232, 41.309724, 5.716078]
|layer_1_from_bottom|back_of_welding_region|left_side_of_welding_region|below_proxy_plane|
tube_lane_
1|proxy_target_0|corner|layer_1_from_bottom,left_side_of_welding_region","5|[19.553238,
39.351494,
29.802536]|layer_2_from_bottom|back_of_welding_region|left_side_of_welding_region|
above_proxy_plane|
tube_lane_1|proxy_target_0|corner|layer_2_from_bottom,left_side_of_welding_region
","6|[10.846296,
49.692977, 20.955645]|layer_2_from_bottom|back_of_welding_region|
left_side_of_welding_region|above_p
roxy_plane|tube_lane_1|proxy_target_0|corner|layer_2_from_bottom,back_of_welding_region,
left_side_of_welding_region","7|[110.698897, 48.278821, 32.800928]
|layer_2_from_bottom|back_of_welding_region|right_side_of_welding_region|above_proxy_plane|
tube_lane
_4|proxy_target_4|corner|layer_2_from_bottom,back_of_welding_region,
right_side_of_welding_region",
"8|[109.98443, 18.736458, 20.649616]|layer_2_from_bottom|front_of_welding_region|
right_side_of_weldi
ng_region|above_proxy_plane|tube_lane_4|proxy_target_4|corner|layer_2_from_bottom,
front_of_welding_region,right_side_of_welding_region","9|[5.375053, 14.823194, 24.447512]
|layer_2_from_bottom|front_of_welding_region|left_side_of_welding_region|above_proxy_plane|
tube_lane
_1|proxy_target_0|corner|layer_2_from_bottom,front_of_welding_region,
left_side_of_welding_region",
"10|[4.525385, 10.917064, 7.610493]|layer_1_from_bottom|front_of_welding_region|
left_side_of_welding
_region|below_proxy_plane|tube_lane_1|proxy_target_0|endpoint|layer_1_from_bottom,
front_of_welding_region,left_side_of_welding_region"],"point_semantics_row_format":"index|
xyz|layer|
front_back|left_right|vertical_relation|nearest_tube_lane|nearest_proxy_target|
local_path_role|group
_ids","control_variable_options":[{"variable_id":"layer_1_from_bottom:x",
"target_type":"point_group","axis":"x","application":"same_delta_to_all_points"},
{"variable_id":"layer_1_from_bottom:y","target_type":"point_group","axis":"y",
"application":"same_delta_to_all_points"},{"variable_id":"layer_1_from_bottom:z",
"target_type":"point_group","axis":"z","application":"same_delta_to_all_points"},
{"variable_id":"layer_2_from_bottom:x","target_type":"point_group","axis":"x",
"application":"same_delta_to_all_points"},{"variable_id":"layer_2_from_bottom:y",
"target_type":"point_group","axis":"y","application":"same_delta_to_all_points"},
{"variable_id":"layer_2_from_bottom:z","target_type":"point_group","axis":"z",
"application":"same_delta_to_all_points"},{"variable_id":"front_of_welding_region:x",
"target_type":"point_group","axis":"x","application":"same_delta_to_all_points"},
{"variable_id":"front_of_welding_region:y","target_type":"point_group","axis":"y",
"application":"same_delta_to_all_points"},{"variable_id":"front_of_welding_region:z",
"target_type":"point_group","axis":"z","application":"same_delta_to_all_points"},
{"variable_id":"back_of_welding_region:x","target_type":"point_group","axis":"x",
"application":"same_delta_to_all_points"},{"variable_id":"back_of_welding_region:y",
"target_type":"point_group","axis":"y","application":"same_delta_to_all_points"},
{"variable_id":"back_of_welding_region:z","target_type":"point_group","axis":"z",
"application":"same_delta_to_all_points"},{"variable_id":"left_side_of_welding_region:x",
"target_type":"point_group","axis":"x","application":"same_delta_to_all_points"},
{"variable_id":"left_side_of_welding_region:y","target_type":"point_group","axis":"y",
"application":"same_delta_to_all_points"},{"variable_id":"left_side_of_welding_region:z",
"target_type":"point_group","axis":"z","application":"same_delta_to_all_points"},
{"variable_id":"right_side_of_welding_region:x","target_type":"point_group","axis":"x",
"application":"same_delta_to_all_points"},{"variable_id":"right_side_of_welding_region:y",
"target_type":"point_group","axis":"y","application":"same_delta_to_all_points"},
{"variable_id":"right_side_of_welding_region:z","target_type":"point_group","axis":"z",
"application":"same_delta_to_all_points"}]}

## Case Memory Index
# Case Memory Index
```

Prompt: Semantic Action Planner *(continued)*

```
## Best So Far
- verified_round_id: 90
- objective_value: 3.65
- metric_profile: {'weld_min_loss': 167.43, 'uniformity_error': 4.35}

## Repeated Verified Observations
- Small-scale y moves toward welding region from non-positive objective (<=0) yield
consistent gains
(rounds 93,94).
- Small-scale y moves toward welding region from best objective (3.65) cause objective drop
 (round
91).
- Small-scale y moves toward welding region from moderate state (1.0-2.5) cause drop if
uniformity
near threshold (rounds 96,97).
- Tube pair 2 is the most frequent weakest-loss pair (rounds 90,9...

## Exploration State
{}

## Channel Evidence
{"status":"compacted","optimization_mode":"uniformity_rebalancing","
actionable_channel_focus":"Tube pair 2 has the minimum lower- and upper-tube loss values,
but all pairwise losses exceed the loss threshold.","global_loss_deficit_summary":{"
weld_min_loss_threshold":115.0,
"weld_min_loss_from_channel_mins":167.4332528246656,"pair_count":4,"
pair_count_below_threshold":0,
"all_pairs_below_threshold":false,"largest_pair_deficit":-52.433253,
"smallest_pair_deficit":-59.140237,"pair_deficit_row_format":"pair_index|pair_loss_sum|
deficit_to_we
ld_min_threshold|is_current_floor","pair_deficit_rows":["1|170.527181|-55.527181|False",
"2|167.433253|-52.433253|True","3|172.234665|-57.234665|False","4|174.140237|-59.140237|
False"],
"interpretation":"The current floor pair identifies the lowest loss channel. Because no
pair is below the loss threshold, it is diagnostic rather than the active global target."},
"weld_min_loss_limiting_pairs":{"lower_tube_min_pair_index":2,
"lower_tube_min_loss":101.64923772970258,"upper_tube_min_pair_index":2,
"upper_tube_min_loss":65.78401509496301,"same_pair_limits_both":true,
"weld_min_loss_from_channel_mins":167.4332528246656},"b_proxy_summary":{"min_proxy_index":
null,
"max_proxy_index":null,"min_b_value":null,"max_b_value":null,"spread":null},
"tube_pair_row_format":"pair_index|lower_tube_loss|upper_tube_loss|pair_loss_sum|
is_lower_tube_min|is
_upper_tube_min|is_pair_sum_min|deficit_to_weld_min_threshold",
"tube_pair_rows":["1|103.88338530535083|66.64379608497397|170.5271813903248|False|False|
False|-55.52
7181","2|101.64923772970258|65.78401509496301|167.4332528246656|True|True|True|-52.433253",
"3|106.35976004450825|65.874905044613 47|172.23466508912173|False|False|False|-57.234665",
"4|105.41839948414889|68.72183741862652|174.14023690277543|False|False|False|-59.140237"],
"b_proxy_row_format":"proxy_index|xyz|b_value|is_b_min|is_b_max","b_proxy_rows":[]}

## DSL Schema
{"proposal_type":"semantic_trust_region","objective_focus":["weld_min_loss","
uniformity_error"],
"causal_hypothesis":"short causal statement grounded in verified metrics","search_policy":"
exploit
| explore | escape_plateau","trust_region":{"semantic_groups":["layer_1_from_bottom",
"layer_2_from_bottom","front_of_welding_region","back_of_welding_region","
left_side_of_welding_region","right_side_of_welding_region"],"axes":["x | y | z"],"
direction":"positive |
negative | either | toward_welding_region | away_from_welding_region","scale":"micro |
small |
medium | large"}}
```

Semantic-Action Program δ_t


```
{
  "proposal_type": "semantic_trust_region",
  "objective_focus": [
    "uniformity_error"
  ],
  "causal_hypothesis": "Tube pair 2 has the lowest lower- and upper-
tube responses. Moving the upper layer rightward should increase
coupling to this relatively weaker pair and reduce inter-pair loss
imbalance.",
  "search_policy": "explore",
  "trust_region": {
    "semantic_groups": [
      "layer_2_from_bottom"
    ],
    "axes": [
      "x"
    ],
    "direction": "positive",
    "scale": "medium"
  }
}
```