STRIDE: STRUCTURAL TOOL-GUIDED REASONING FOR INTELLIGENT DESIGN AND EVALUATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Computable beam-and-column framing layouts encode load paths for early structural analysis, member sizing, and architectural design exploration. Constructing these layouts from architectural floor plans remains difficult because column locations, beam connectivity, and member sizes interact through geometry and structural response. An effective design procedure revises candidate graphs using computed responses while managing the cost of structural analysis and member sizing. We introduce STRIDE (Structural Tool-Guided Reasoning for Intelligent Design and Evaluation), an LLM agent that reads a plan raster and wall graph, generates a beam-and-column framing graph, and uses finite-element feedback to revise it. It links governing member responses to editable beam and column decisions, uses differentiable frame sensitivities to size members, and reevaluates proposed graphs, closing the loop from plan interpretation to structural revision. Across six real architectural projects, STRIDE generates layouts that pass structural response checks; its feedback-guided revisions reduce gross beam-and-column volume by 13.95% on average under a common member-sizing budget. The resulting framing graphs retain explicit beam connectivity and column locations, compile into three-dimensional finite-element models, and support further structural design iterations.

1 INTRODUCTION

An architectural plan organizes spaces and boundaries; a structural graph connects beams and columns to carry loads to the ground. Graph edits change member spans and forces, while subsequent section sizing changes their cost.

STRIDE maps finite-element responses to graph edits and resizes each proposal before comparing it with its parent.

STRIDE implements this interface on real architectural inputs. A multimodal initializer receives the plan raster and metric wall graph and produces one typical-floor framing graph. A typed editor can move joints, add or remove columns, split or remove beams, and reconnect incidence. A sparse differentiable frame solver supplies section sensitivities, while explicit checks of selected catalog sizes determine acceptance. Selected sections, responses, and adjacency are mapped back to the editable graph for the next proposal.

We study six architectural projects—U-plan, Sloped-edge, Radial, Winged, Stepped, and Notched—ranging from two to fourteen stories. Figure 1 shows the agent-to-mechanics loop; Appendix Figures 4 and 5 show the architectural inputs and generated structural graphs.

We compare feedback-guided revisions with continued parent sizing and deterministic controls, measuring member volume, column locations, column area, and beam proximity. In the U-plan sequence in Figure 1, three revisions reduce gross member volume from 429.98 to 226.46 m^3.

2 RELATED WORK

Architectural graphs and structural generation. Graph-based methods generate beam layouts and structural topology from architectural organization (Zhao et al., 2023b; Zhang et al., 2024). To-

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.

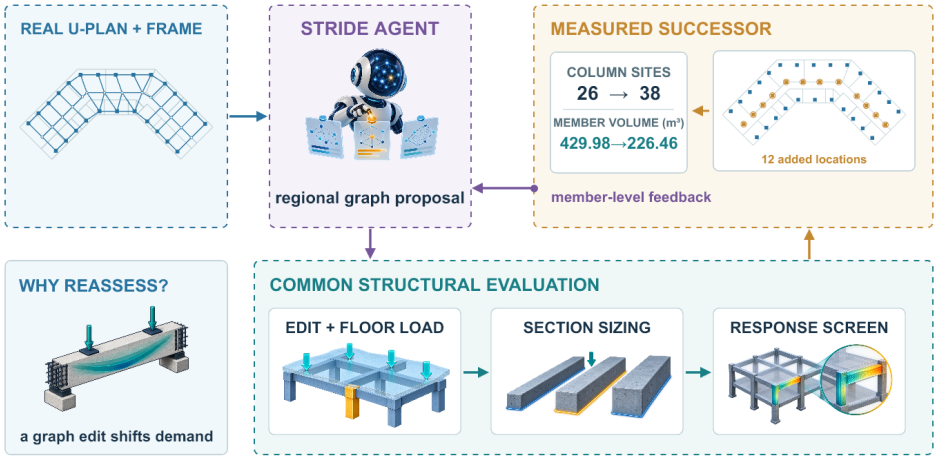

Figure 1: STRIDE's U-plan graph-edit loop. Orange marks added column locations; the avatar and engineering cutouts visualize the workflow.

poGNN trains separate column and beam prediction modules on real architectural drawing graphs and integrates their outputs into a frame topology (Long et al., 2026). Trained generative models extend learned structural generation to wall systems, including GAN-based shear-wall design, physics-enhanced generation, graph- and channel-based wall representations, diffusion models, and wall–beam co-design (Liao et al., 2021; Lu et al., 2022; Zhao et al., 2023c; Han et al., 2024; Gu et al., 2024; Xia et al., 2025). Wall-structure extensions cover beam layouts, design-condition-informed generation, beam-and-slab deep learning, knowledge-enhanced tube GANs, GNN-assisted evolutionary optimization, and diffusion generation with analysis (Zhao et al., 2024; 2023a; 2022; Fei et al., 2022; 2025b; Zhou et al., 2024), with data enhancement and a field review in (Fei et al., 2025a; Liao et al., 2024). These approaches build on graph convolutional networks (Kipf & Welling, 2017) and graph-network simulators of physical systems (Pfaff et al., 2021). These learned generators are trained for their target design tasks. STRIDE drives a frozen general-purpose model with member-linked numerical feedback, uses an explicit graph-edit grammar, and tests how each graph decision changes after finite catalog sizing. Optimization of column layouts under architectural and structural constraints is also established (Zelickman & Amir, 2026). We report column-location count, summed column area, and member volume separately.

Agents with engineering tools. Transformer-based general-purpose models (Vaswani et al., 2017) act as design agents: ReAct interleaves reasoning with actions and observations (Yao et al., 2023); Toolformer learns tool calls from self-supervision (Schick et al., 2023); Reflexion uses verbal reflection (Shinn et al., 2023), and Self-Refine iterates on self-feedback (Madaan et al., 2023). In structural design, an LLM controller has translated engineer instructions into calls to shear-wall generation and topology, pattern, and size optimization modules (Qin et al., 2024a). STRIDE asks the model to choose typed edits to a beam-and-column graph from member-linked finite-element responses, then compares those edits with continued sizing and deterministic graph controls at declared numerical budgets. Other engineering applications include conceptual topology optimization (Liang et al., 2025), structural-software operation through a knowledge graph (Chen & Bao, 2026), a generative-AI structural design copilot (Qin et al., 2024b), and agentic frameworks that carry architectural sketches to fabrication-ready truss documentation (Meng et al., 2026). Agent systems have also been applied to preference-guided and self-refining continuum topology optimization through multi-agent frameworks (Stewart et al., 2026; Park & Chung, 2026).

Finite inner optimization. Discrete section optimization can itself be learned or searched (Hayashi & Ohsaki, 2022). More generally, inexact lower-level solutions affect bilevel decisions (Diouane et al., 2024). Multi-fidelity and constrained Bayesian optimization treat expensive model evaluations as a budget (Tran et al., 2020; Paulson & Lu, 2022), and machine-learning accelera-

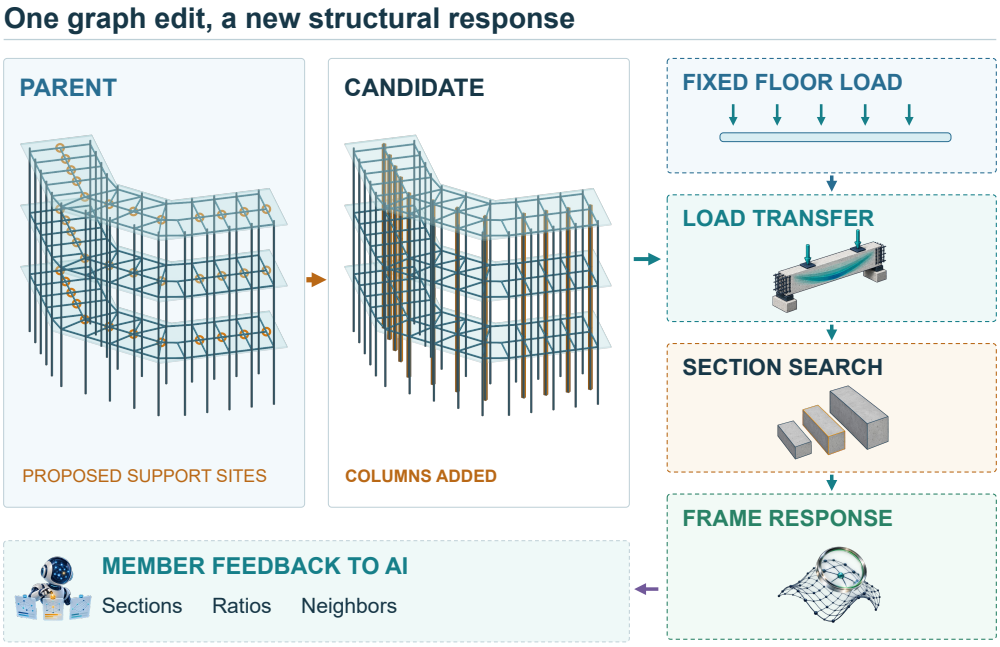

Figure 2: U-plan graph edit and member-linked structural feedback. The frame-response cutout illustrates the feedback interface.

tion speeds structural optimization (Xing & Tong, 2022). STRIDE evaluates graphs through finite catalog sizing and compares each edit with continued sizing of its unchanged parent.

3 Task and Method

3.1 Architectural Input and Structural Actions

Input $A = (I, W, \Omega, H)$ consists of a plan raster I, a metric wall graph W, the architectural outer boundary Ω, and the supplied story count H. The structural template $T = (X, E, c)$ contains planar joint coordinates X, beam incidence E, and column-presence indicators $c_i \in \{0, 1\}$. Expansion repeats the template at the declared levels and instantiates a continuous column at each selected plan location. Plan coordinates use a 50 mm grid; every crossing or T-junction must be represented by shared incidence.

Each floor has one continuous gravity-load surface enclosed by Ω. The full enclosed area carries floor pressure, including space inside internal wall rings and at column locations. Column footprints are checked separately against Ω. Nodes and beam centerlines must lie inside the boundary, and every beam component must connect to a column load path.

A proposal consists of at most four ordered actions: move a joint, toggle its column status, add or remove a beam, split a beam with a new joint, or remove a joint with an admissible reconnection. The batch is atomic. Invalid incidence, crossings, or spatial geometry rejects the proposal and consumes its slot. Source wall directions and connected wall groups define local frames (o_r, R_r), with $x = o_r + R_r[u, v]^\top$ for coordinate-valued actions. The canonical global frame is always available in the full method.

3.2 DIFFERENTIABLE MECHANICS AND FEASIBLE CATALOG SIZING

For framing graph T and grouped cross-section assignment s (the section schedule), the shared linear-elastic frame model solves

$$K(T, s)u_\ell = f_\ell(T, s), \qquad \ell \in \mathcal{L}, \tag{1}$$

over five serviceability and five factored-load combinations. Sparse assembly and a shared factorization handle the ten right-hand sides. Implicit differentiation reuses this factorization for section sensitivities. The load vector includes the complete floor pressure and geometry-dependent beam and column selfweight. Tributary allocation distributes floor pressure to the current beam graph while preserving the imposed floor resultant. The response screen evaluates linear-elastic three-dimensional frame behavior under the declared full-floor gravity and equivalent-static wind loads. The wind-force resultant is fixed per project, while its equal allocation to joints at selected column locations follows the edited layout; the application centroid and in-plane moment can therefore change (Appendix A).

The sizing routine returns the lowest-volume screen-passing catalog schedule visited within budget B:

$$\widehat{V}_B(T) = \min_{s \in \mathcal{S}_B(T):\, g_\ell(T,s) \le 0 \ \forall \ell} \sum_{e \in E_{3D}(T)} b_e h_e L_e. \tag{2}$$

Here $\mathcal{S}_B(T)$ is the finite assessed schedule set. For graphs with a returned feasible schedule, write $\widehat{V}_B(T) = V^\star(T) + \epsilon_B(T)$, where $V^\star$ is the minimum over all feasible grouped schedules in the same section catalog and $\epsilon_B \ge 0$ is unobserved finite-search error. Thus an observed graph-edit gain decomposes as

$$\widehat{V}_B(T) - \widehat{V}_B(T') = [V^\star(T) - V^\star(T')] + [\epsilon_B(T) - \epsilon_B(T')]. \tag{3}$$

The controls below test how observed rankings change with sizing effort. The service begins with the largest geometrically admissible catalog sections. At each descent state it differentiates a fourth-power aggregation of dimensionless response ratios, ranks one-entry catalog reductions by material saving relative to predicted response increase, and proposes grouped reductions. Repeated-floor beams and three-story column bands share decisions. A hard assessment accepts a batch only when all declared limits pass and volume falls; otherwise the batch is reduced. Both sensitivity states and hard trial schedules consume B. A distinct infeasible-start status records cases in which the largest admissible schedule exceeds the response screen.

3.3 CONNECTED FEEDBACK AND THE OUTER LOOP

The selected schedule is mapped back from three-dimensional members to template members. For each member, the report retains governing load cases, response ratios, selected sections, admissible catalog entries, and adjacent joints and beams. The compact regional view includes twelve critical neighborhoods, eight low-utilization column neighborhoods, and eight material hotspots. The governing- witness view contains the same objective, global checks, and worst-member witnesses. Both receive identical architectural walls, regions, interfaces, section catalogs, and floor geometry.

Given feedback F_t and recorded edit history M_t, the agent proposes a batch $a_t = \pi(A, T_t, F_t, M_t)$. The candidate $T'_t = a_t(T_t)$ passes geometric checks and receives a fresh section assessment. A feasible lower-volume candidate becomes the incumbent; the incumbent persists for every other candidate. History records the proposed actions and their assessed outcome. Maximum and median distances from a fixed 1 m floor-sample grid to the nearest beam centerline may increase by at most 0.05 m from the initial parent. This beam-proximity guard preserves the initial framing distribution; the absolute distances are also reported when comparing independent constructions.

4 EXPERIMENTAL DESIGN

Projects and engineering configuration. The cohort comprises six owner-edited architectural graph-and-raster pairs. All generated and comparison templates use the same source-derived outer boundaries and repeated standard floors. Prompts and baseline construction use architectural inputs alone. We use a common engineering configuration: 3 m story height, 30 GPa elastic modulus,

concrete density 25 kN/m^3, 0.12 m floor thickness, 3 kPa slab dead pressure, 2 kPa additional dead pressure, 2 kPa live pressure, and signed 0.5 kPa wind pressure in each horizontal direction. Beam sections range from 0.20×0.40 to 0.40×1.00 m; square column sides range from 0.40 to 1.20 m. Appendix A specifies combinations, screening equations, and catalog entries.

Initialization and construction controls. One request configured for GPT-6 Sol/high receives the raster and architectural graph for each project and returns a structural template. We also assess 31 source-geometric constructions, including rotated grids, wall networks, a radial construction, and perimeter-preserving networks. We report the assessment status of all 37 model and geometric templates. Graph-edit comparisons begin from common generated parents.

Member-sizing comparison. We assess each generated graph and three U-plan edited graphs with both straight-through catalog optimization and feasible descent under the same full-floor gravity-load model and a cap of 81 sizing factorizations.

Paired decisions and trajectories. For each project, a 2×2 decision screen crosses regional versus governing- witness feedback with global-only versus global-plus-local action coordinates. Each cell receives one new model request from the identical assessed parent. A source-aware deterministic editor receives the same full feedback, source geometry, and action grammar. It considers column removal, material-driven column addition, beam splits/removals, and geometry-guided repairs, screening its candidates from supplied feedback before selecting one proposal. We then issue three sequential requests per project for regional feedback, governing-witness feedback, and the deterministic editor, all with local-admissible coordinates. These methods were set before inspecting the paired-decision results. A broader deterministic editor uses the same three proposal slots, action limit, coverage guard, regional feedback, and sizing budget. It cycles column addition at high-volume beam ends, removal of low-utilization columns, beam splits, and column relocation. Two further rules test repeated hotspot promotion and adaptive ranking across five edit families. Both operate deterministically on supplied feedback. The appendix specifies their rules and available endpoint checks.

Every assessed topology uses at most 81 sizing factorizations plus one initial and one feedback factorization, each covering ten load combinations. Every proposal slot counts as a query. We report per-project feasible volume improvement, action and assessment validity, column-location count and summed column cross-sectional area, beam length, and floor-point-to-beam distances. Projects remain the paired units; floors, load combinations, and repeated requests remain observations within each project.

5 Results

5.1 From architectural inputs to assessed structural graphs

The first request for each project returns a valid structural template: all pass graph incidence, spatial containment, fitting-column, and supported-component checks. The generated graphs contain 31–71 joints, 46–118 beams, and 26–71 column locations per typical floor. Common feasible descent returns a feasible section schedule for every project under the declared frame response screen across ten load cases (Table 1). Independent OpenSees (Zhu et al., 2018) checks agree with the sparse solver for all 60 project–load combinations and retain every feasibility label. The highest independent screen ratio is 0.99868; Appendix A gives the agreement discrepancies and tolerances.

Across the model starts and 31 source-geometric constructions, all 37 templates pass geometry checks and 28 return feasible section schedules. The lowest-volume feasible geometric seed in each project uses less material than the one-shot model graph, but has a larger 95th-percentile floor-point-to-beam distance. Five model starts are nondominated by feasible constructions in material volume and 95th-percentile floor-point-to-beam distance; geometric construction sets the two-metric frontier for Stepped. Feedback comparisons use common generated parents; Appendix tables report all candidates.

Table 1: Generated templates and section-sizing comparison for real projects. H is story count; C and E count column locations and beams per floor. Volumes are total gross member m^3. Both algorithms use 81 sizing factorizations. † marks a response-screen exceedance.

Project	H	C	E	Straight-through	Feasible descent
U-plan	3	26	66	$632.12^{\dagger}$	429.98
Sloped-edge	6	39	75	$1054.86^{\dagger}$	813.63
Radial	2	43	98	$485.95^{\dagger}$	289.49
Winged	4	71	118	$793.28^{\dagger}$	1251.90
Stepped	9	29	46	$1033.33^{\dagger}$	785.86
Notched	14	44	69	2528.38	2533.36

Table 2: Selected feasible volume after three proposals from each common parent. Entries show m^3 and reduction (%). Regional denotes regional feedback; Witness, governing-witness feedback; Deterministic, the prespecified source-aware editor; and Adaptive rule, the post-hoc adaptive five-family editor. Each method has one trajectory per project and eighteen proposal slots. Both model editors use local-admissible coordinates.

Project	Regional	Witness	Deterministic	Adaptive rule
U-plan	226.46 (47.33)	382.05 (11.15)	429.98 (0.00)	226.46 (47.33)
Sloped-edge	769.74 (5.40)	751.17 (7.68)	805.95 (0.94)	801.97 (1.43)
Radial	268.97 (7.09)	270.41 (6.59)	278.76 (3.70)	288.20 (0.44)
Winged	1166.21 (6.85)	1144.37 (8.59)	1169.82 (6.56)	1172.55 (6.34)
Stepped	698.61 (11.10)	700.52 (10.86)	785.84 (0.00)	785.86 (0.00)
Notched	2382.65 (5.95)	2218.21 (12.44)	2415.58 (4.65)	2276.08 (10.16)
Mean reduction	13.95%	9.55%	2.64%	10.95%

5.2 Three-query topology trajectories

Both model editors improve the selected feasible design on every project (Table 2). Regional feedback achieves a 13.95% mean per-project reduction, governing-witness feedback 9.55%, and the prespecified source-aware deterministic editor 2.64%. The mean gap is driven by U-plan (47.33% regional versus 11.15% witness); across the other five projects, mean reductions are 7.28% and 9.23%, respectively. Regional feedback yields lower volume in U-plan, Radial, and Stepped; witness feedback yields lower volume in the other three projects. The post-hoc adaptive five-family editor reaches a 10.95% mean under the same three-slot and per-topology sizing limits. At this primary cap, regional editing has lower member volume on four projects, ties on U-plan, and has higher volume on Notched.

What the U-plan improvement changes. The regional trajectory adds column locations at twelve existing beam-only joints, increasing column locations from 26 to 38 while preserving beam incidence and floor coverage. The graph remains on the same full-floor gravity-load domain. Its volume falls by 203.52 m^3 after resizing: beam volume changes from 324.23 to 170.12 m^3, and column volume from 105.75 to 56.34 m^3. Across three stories, 36 added column segments coincide with a reduction in summed column cross-sectional area from 11.75 to 6.26 m^2 per floor. Column-location count and cross-sectional area therefore move in opposite directions; Appendix Figure 6 locates the added columns against the source walls. Eleven of the twelve added column locations lie within 0.10 m of source walls. For a re-ranking objective $V + \lambda_{\mathrm{seg}} N_{\mathrm{seg}}$, the break-even penalty is 5.65 m^3-equivalent per added segment. A column-location-count objective instead has a 16.96 m^3-equivalent break-even penalty per added location. These are modeled member-volume thresholds.

5.3 Inner sizing changes the feedback supplied to topology editing

Under identical full-floor gravity loads, straight-through catalog optimization from the declared baseline sections finds a feasible schedule for one generated graph. Feasible descent from the largest

Table 3: One-query reduction in feasible member volume from identical parents (%). Local admits both source-local and global frames. Each model cell has one request per project; the deterministic column has one source-aware proposal per project.

	Regional local	Regional global	Witness local	Witness global	Deterministic
U-plan	21.50	21.50	0.00	11.92	0.00
Sloped-edge	0.00	0.00	6.85	6.74	0.00
Radial	0.00	0.00	3.47	0.00	3.70
Winged	2.51	3.18	2.05	3.10	1.99
Stepped	8.97	8.97	0.00	8.97	0.00
Notched	0.00	0.00	8.81	0.00	4.65
Mean	5.50	5.61	3.53	5.12	1.72

admissible sections finds one for every project at the same factorization cap. On Notched, both procedures succeed and straight-through returns 4.98 m^3 less material.

On U-plan, three successive graph states add four, eight, and twelve column locations to the unchanged parent with 26 column locations. Feasible descent returns passing schedules at 358.14, 317.35, and 246.20 m^3, respectively, compared with the parent's 429.98 m^3. The added columns reduce modeled member volume. Straight-through sizing exceeds the response screen on all three edited states.

5.4 Paired feedback and coordinate interventions

All 24 model responses parse. Nineteen proposals pass the action engine, eighteen reach numerical assessment after the coverage check, and fourteen improve their common-parent incumbent. Table 3 includes rejected proposals as zero improvement. With local frames available, the regional-minus-witness difference in mean volume reduction is 1.97 percentage points. The per-project difference is positive in three projects and negative in three, with a median of -1.51 points; with global-only actions, the mean difference is 0.49 percentage points.

All 60 realized paired and sequential actions used coordinate-free edit types. Appendix A.7 reports request tokens and mechanics calls.

5.5 Repeated proposals and longer section search

Independent requests repeat both model policies on the U-plan (largest gain), Radial, and Notched (tallest), keeping the same parents and three-query limits. Regional reductions are 47.33%, 3.14%, and 6.45%; witness reductions are 4.19%, 7.68%, and 8.81%, respectively. The large U-plan regional improvement recurs, while the other two projects favor the witness policy in this repeat. Separate later samples on the remaining three projects favor witnesses in each pair (Appendix Table 14).

The broader deterministic editor improves four initial designs at the 81-factorization cap, with a 7.04% project-level mean reduction. Each model policy exceeds it on five projects. At the 321-factorization cap, U-plan's deterministic endpoint retains a 23.70% reduction relative to its unchanged parent; two projects tie and three favor the parent. All selected schedules pass independent ten-load-case response checks.

Persistent hotspot promotion reproduces U-plan's regional graph and selected sections at 226.46 m^3, then retains the other five parents. The post-hoc adaptive five-family editor improves five parents and averages 10.95% reduction, compared with regional editing's 13.95% and witnesses' 9.55%. Appendix Table 7 reports every project, proposal count, and factorization total. Independent OpenSees checks pass all distinct adaptive-editor parent and endpoint states at both the 81- and 321-factorization caps (220 load-case runs). At the larger cap, its U-plan, Sloped-edge, and Radial graphs remain lighter than their newly sized parents; Appendix Table 8 gives all paired results. At the same 321-factorization cap, the adaptive editor ties the regional graph on U-plan and is lighter

		U-plan	Sloped-edge	Radial	Winged	Stepped	Notched
Primary 81 fact.	Regional	+47.3	+5.4	+7.1	+6.8	+11.1	+5.9
	Witness	+11.1	+7.7	+6.6	+8.6	+10.9	+12.4
	Mechanics	0.0	+0.9	+3.7	+6.6	+<0.1	+4.6
Extended 321 fact.	Regional	+49.7	+1.1	-1.5	-19.4	-16.3	-71.2
	Witness	+13.5	-0.7	-3.2	-32.3	+2.5	-8.4
	Mechanics	0.0	-29.8	-7.0	-35.0	-31.7	-46.4

Figure 3: Frozen graphs assessed under primary sizing (81 factorizations) and extended sizing (321 factorizations). Each cell gives the signed percent member-volume saving against its same-project, same-budget unchanged parent: positive/blue means less material and negative/amber means more. The groups are separate observed evaluations. Mechanics is the prespecified source-aware deterministic editor.

on the other five projects. Its mean parent-relative reduction is 5.05%, versus -9.59% for regional editing. This comparison gives each fixed graph the same sizing cap; a separate experiment below matches total trajectory calls. The lower-volume adaptive editor uses more column locations than the regional graph on Sloped-edge, Radial, Winged, and Stepped; the two graphs tie on U-plan, while the adaptive editor uses fewer column locations on Notched. Regional summed column section area is lower on Sloped-edge, Radial, and Stepped. Appendix Table 9 reports the selected sections and column-location counts at the same sizing cap. The four resulting member-volume–column-location tradeoffs have break-even penalties from 0.102 to 29.457 m^3-equivalent per column location for $V + \lambda_{\text{col}} N_{\text{col}}$ (Appendix Table 10). Across the five common-budget method states per project, the regional endpoint is Pareto-nondominated in member volume and column-location count in five projects (Appendix Table 11).

We resize every unchanged parent and original method endpoint from 81 to 321 factorizations. The 24 parent and endpoint slots contain 23 distinct graphs because U-plan's deterministic endpoint is unchanged. Independent OpenSees checks pass for all 23 graphs at both section budgets, covering 460 state–load comparisons. Longer section search changes the material ranking (Figure 3). U-plan's regional endpoint retains a 49.70% reduction relative to its newly sized parent. Regional and witness improvements each persist on two projects; the deterministic endpoints tie or favor their newly sized parents. The project-level mean changes are -9.59%, -4.78%, and -24.97%, respectively, when every edited endpoint is compared to its own newly sized parent. Negative values indicate lower volume for the newly sized parent.

At 4,097 factorizations, all 18 parent–endpoint rankings remain as at 321; 13 differ from the 81-factorization assessment. U-plan's regional graph remains 49.82% lighter than its parent. Appendix Table 13 reports every endpoint and stopping status.

At matched total FE factorizations (at most 332 per trajectory), the two U-plan model edits beat continued sizing of their unchanged parent; one control ties, and continued parent sizing yields lower volume in the other 15 comparisons. Appendix Table 12 reports each matched comparison.

6 Discussion and Conclusion

STRIDE connects plan-conditioned graph proposals, differentiable sizing, and member-linked finite-element feedback. Regional trajectories reduce member volume on all six projects at the primary sizing cap (13.95% mean). Continued sizing and common-budget comparisons show how section-search effort changes graph rankings. U-plan retains a 49.82% reduction at 4,097 factorizations under the declared wind-force allocation; summed column area also falls across twelve added locations. The workflow exposes member-volume, column-location, and frame-response tradeoffs for each proposed graph.

REPRODUCIBILITY STATEMENT

The appendix and anonymous supplement provide the numerical protocol, selected code, project graphs, saved prompts and responses, and structural results.

AI USE STATEMENT

Generative AI assisted study design, coding, analysis, literature search, figures, and writing. An LLM proposed framing graphs in the experiments. The authors take responsibility for the work.

REFERENCES

Jinxin Chen and Yi Bao. Artificial intelligence design agent for automated operation of structural software using knowledge graphs and numerical models. *Automation in Construction*, 190: 107144, 2026. doi: 10.1016/j.autcon.2026.107144. URL https://www.sciencedirect.com/science/article/pii/S0926580526003857.

Youssef Diouane, Vyacheslav Kungurtsev, Francesco Rinaldi, and Damiano Zeffiro. Inexact direct-search methods for bilevel optimization problems. *Computational Optimization and Applications*, 88:469–490, 2024. doi: 10.1007/s10589-024-00567-7. URL https://link.springer.com/article/10.1007/s10589-024-00567-7.

Yifan Fei, Wenjie Liao, Yuli Huang, and Xinzheng Lu. Knowledge-enhanced generative adversarial networks for schematic design of framed tube structures. *Automation in Construction*, 144: 104619, 2022. doi: 10.1016/j.autcon.2022.104619.

Yifan Fei, Xinzheng Lu, Wenjie Liao, and Hong Guan. Data enhancement for generative ai design of shear wall structures incorporating structural optimization and diffusion models. *Advances in Structural Engineering*, 29:545–565, 2025a. doi: 10.1177/13694332251353614.

Yifan Fei, Sizhong Qin, Wenjie Liao, Hong Guan, and Xinzheng Lu. Graph neural network-assisted evolutionary algorithm for rapid optimization design of shear-wall structures. *Advanced Engineering Informatics*, 65:103129, 2025b. doi: 10.1016/j.aei.2025.103129.

Yi Gu, Yuli Huang, Wenjie Liao, and Xinzheng Lu. Intelligent design of shear wall layout based on diffusion models. *Computer-Aided Civil and Infrastructure Engineering*, 39:3610–3625, 2024. doi: 10.1111/mice.13236.

Jin Han, Xinzheng Lu, Yi Gu, Wenjie Liao, Qing Cai, and Hongjing Xue. Optimized data representation and understanding method for the intelligent design of shear wall structures. *Engineering Structures*, 315:118500, 2024. doi: 10.1016/j.engstruct.2024.118500.

Kazuki Hayashi and Makoto Ohsaki. Graph-based reinforcement learning for discrete cross-section optimization of planar steel frames. *Advanced Engineering Informatics*, 51:101512, 2022. doi: 10.1016/j.aei.2021.101512. URL https://www.sciencedirect.com/science/article/pii/S1474034621002603.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, 2017. doi: 10.48550/arXiv.1609.02907. URL https://arxiv.org/abs/1609.02907.

Zelong Liang, Yuan-Fang Zhang, Yingjun Wang, and Weihua Li. Integrating large models with topology optimization for conceptual design realization. *Advanced Engineering Informatics*, 67: 103524, 2025. doi: 10.1016/j.aei.2025.103524. URL https://www.sciencedirect.com/science/article/pii/S1474034625004173.

Wenjie Liao, Xinzheng Lu, Yuli Huang, Zhe Zheng, and Yuanqing Lin. Automated structural design of shear wall residential buildings using generative adversarial network. *Automation in Construction*, 132:103931, 2021. doi: 10.1016/j.autcon.2021.103931.

Wenjie Liao, Xinzheng Lu, Yifan Fei, Yi Gu, and Yuli Huang. Generative ai design for building structures. *Automation in Construction*, 157:105187, 2024. doi: 10.1016/j.autcon.2023.105187.

Danbing Long, Yongjun Kang, Guanjun Liu, Changjian Fang, Guangpo Zhao, and Yifeng Lai. Two-stage automated generation of frame structure topology using graph neural networks. *Automation in Construction*, 181:106627, 2026. doi: 10.1016/j.autcon.2025.106627. URL https://www.sciencedirect.com/science/article/pii/S0926580525006673.

Xinzheng Lu, Wenjie Liao, Yu Zhang, and Yuli Huang. Intelligent structural design of shear wall residence using physics-enhanced generative adversarial network. *Earthquake Engineering & Structural Dynamics*, 51:1657–1676, 2022. doi: 10.1002/eqe.3632.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-Refine: Iterative Refinement with Self-Feedback. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2019. URL https://papers.nips.cc/paper/2023/hash/91edff07232fb1b55a505a9e9f6c0ff3-Abstract-Conference.html.

Xianchuan Meng, Ziyu Tong, and Chen Zhao. From architectural sketches to fabrication-ready documentation: Agentic framework for planar truss structures. *Automation in Construction*, 188: 107010, 2026. doi: 10.1016/j.autcon.2026.107010.

Hyunjee Park and Hayoung Chung. Self-refining topology optimization via an LLM-based multi-agent framework, 2026. URL https://arxiv.org/abs/2605.23273.

Joel A. Paulson and Congwen Lu. Cobalt: Constrained bayesian optimization of computationally expensive grey-box models exploiting derivative information. *Computers & Chemical Engineering*, 160:107700, 2022. doi: 10.1016/j.compchemeng.2022.107700.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning Mesh-Based Simulation with Graph Networks. In *International Conference on Learning Representations*, 2021. doi: 10.48550/arXiv.2010.03409. URL https://arxiv.org/abs/2010.03409.

Sizhong Qin, Hong Guan, Wenjie Liao, Yi Gu, Zhe Zheng, Hongjing Xue, and Xinzheng Lu. Intelligent design and optimization system for shear wall structures based on large language models and generative artificial intelligence. *Journal of Building Engineering*, 95:109996, 2024a. doi: 10.1016/j.jobe.2024.109996. URL https://www.sciencedirect.com/science/article/abs/pii/S235271022401564X.

Sizhong Qin, Wenjie Liao, Shengnan Huang, Kongguo Hu, Zhuang Tan, Yuan Gao, and Xinzheng Lu. Aistructure-copilot: Assistant for generative ai-driven intelligent design of building structures. *Smart Construction*, 2024b. doi: 10.55092/sc20240001.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2997.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language Agents with Verbal Reinforcement Learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-0377. URL https://papers.nips.cc/paper_files/paper/2023/hash/1b44b878bb782e6954cd888628510e90-Abstract-Conference.html.

Isabella A. Stewart, Hongrui Chen, and Faez Ahmed. TO-Agents: A Multi-Agent AI Framework for Subjective Preference-Guided Topology Optimization, 2026. URL https://arxiv.org/abs/2605.21622v2. Accepted for publication in the Proceedings of the ASME 2026 International Design Engineering Technical Conferences (IDETC2026).

Anh Tran, Tim Wildey, and Scott McCann. smf-bo-2cogp: A sequential multi-fidelity constrained bayesian optimization framework for design applications. *Journal of Computing and Information Science in Engineering*, 20:031007, 2020. doi: 10.1115/1.4046697.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. doi: 10.48550/arXiv.1706.03762. URL https://arxiv.org/abs/1706.03762.

Jikang Xia, Wenjie Liao, Bo Han, Shulu Zhang, and Xinzheng Lu. Intelligent co-design of shear wall and beam layouts using a graph neural network. *Automation in Construction*, 172:106024, 2025. doi: 10.1016/j.autcon.2025.106024.

Yi Xing and Liyong Tong. A machine learning-assisted structural optimization scheme for fast-tracking topology optimization. *Structural and Multidisciplinary Optimization*, 65:105, 2022. doi: 10.1007/s00158-022-03181-5.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing Reasoning and Acting in Language Models. In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2210.03629. URL https://arxiv.org/abs/2210.03629.

Yakov Zelickman and Oded Amir. Optimization of column layouts in buildings considering structural and architectural constraints. *Architecture, Structures and Construction*, 6:13, 2026. doi: 10.1007/s44150-026-00196-0. URL https://link.springer.com/article/10.1007/s44150-026-00196-0.

Chong Zhang, Mu-Xuan Tao, Chen Wang, and Jian-Sheng Fan. End-to-end generation of structural topology for complex architectural layouts with graph neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 39(5):756–775, 2024. doi: 10.1111/mice.13098. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.13098.

Pengju Zhao, Wenjie Liao, Hongjing Xue, and Xinzheng Lu. Intelligent design method for beam and slab of shear wall structure based on deep learning. *Journal of Building Engineering*, 57:104838, 2022. doi: 10.1016/j.jobe.2022.104838.

Pengju Zhao, Yifan Fei, Yuli Huang, Yitian Feng, Wenjie Liao, and Xinzheng Lu. Design-condition-informed shear wall layout design based on graph neural networks. *Advanced Engineering Informatics*, 58:102190, 2023a. doi: 10.1016/j.aei.2023.102190.

Pengju Zhao, Wenjie Liao, Yuli Huang, and Xinzheng Lu. Intelligent beam layout design for frame structure based on graph neural networks. *Journal of Building Engineering*, 63:105499, 2023b. doi: 10.1016/j.jobe.2022.105499. URL https://www.sciencedirect.com/science/article/pii/S2352710222015054.

Pengju Zhao, Wenjie Liao, Yuli Huang, and Xinzheng Lu. Intelligent design of shear wall layout based on graph neural networks. *Advanced Engineering Informatics*, 55:101886, 2023c. doi: 10.1016/j.aei.2023.101886.

Pengju Zhao, Wenjie Liao, Yuli Huang, and Xinzheng Lu. Beam layout design of shear wall structures based on graph neural networks. *Automation in Construction*, 158:105223, 2024. doi: 10.1016/j.autcon.2023.105223.

Ying Zhou, Hao Leng, Shiqiao Meng, Hao Wu, and Zheng Zhang. Structdiffusion: End-to-end intelligent shear wall structure layout generation and analysis using diffusion model. *Engineering Structures*, 309:118068, 2024. doi: 10.1016/j.engstruct.2024.118068.

Minjie Zhu, Frank McKenna, and Michael H. Scott. Openseespy: Python library for the opensees finite element framework. *SoftwareX*, 2018. doi: 10.1016/j.softx.2017.10.009.

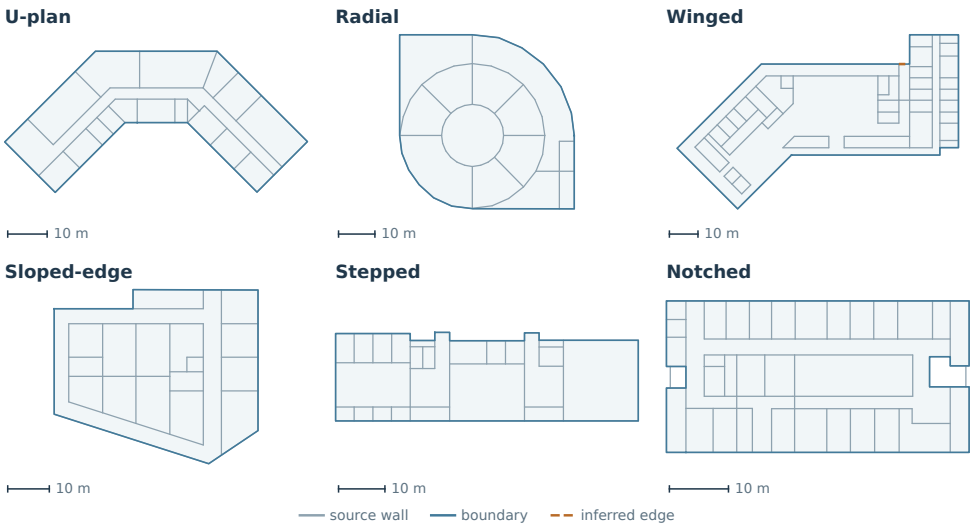

Figure 4: Architectural wall graphs and declared floor outlines before structural generation. The dashed Winged connection marks the inferred boundary closure. Panel scales differ; each scale bar is 10 m.

A Mechanics and Reproduction Details

A.1 Project surfaces and source records

The architectural wall graphs supply geometry; generation uses these records independently of their reference structural graphs. One primary floor surface is repeated at the supplied story count. Table 4 reports the primary computational areas. All selected surfaces are solid polygons, including the interior of Radial's inner wall ring. Winged's outer boundary includes a scenario-declared, raster-supported inferred 3.75 m upper connection.

Table 4. Current owner-confirmed story counts and primary declared floor areas.

Project	Stories	Area (m^2)	Boundary construction
U-plan	3	1,463.035	Closed wall perimeter
Sloped-edge	6	1,693.774	Wall-trace closure
Radial	2	2,033.054	Outer wall ring
Winged	4	3,211.935	Raster-supported wing union
Stepped	9	1,168.165	Stepped outer wall trace
Notched	14	1,273.308	Notched outer wall trace

U-plan and Radial use closed outer rings; Sloped-edge closes its open top trace with three source-wall segments. Stepped follows seven source-wall segments along its stepped edge, and four source-wall segments close Notched's side notches.

Write the one-floor template as $T = (X, E, c)$: X contains planar joint coordinates, E is beam incidence, and $c_i \in \{0, 1\}$ indicates whether joint i is a column location. Coordinates lie on a 50 mm grid. Expansion repeats the template at 3 m story increments and creates a continuous column at each selected plan location, from a six-degree-of-freedom fixed base to the top level. The same plan template is used at every level. Nodes, beam centerlines, and complete square column footprints must fit inside the declared surface; every beam component must connect to a joint with a column. Crossing and T-junctions require shared graph incidence. Beam proximity is measured on a 1 m grid across the full floor as distance to the nearest beam centerline.

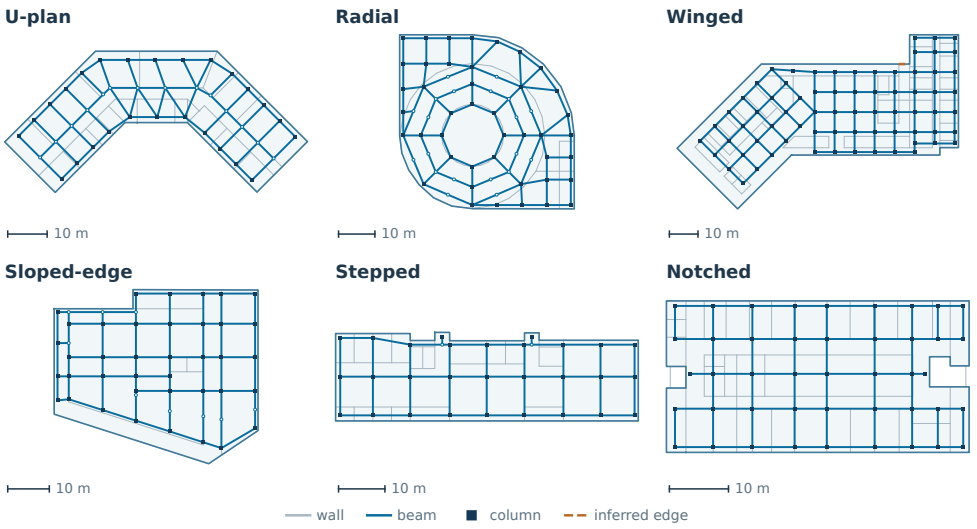

Figure 5: First generated structural graphs. Gray lines are source walls; blue lines and squares are generated beams and column sites. Shading marks the complete floor domain; the dashed orange Winged segment marks its inferred boundary closure. Panel scales differ; each scale bar is 10 m.

A.2 Elastic frame, loads, and limit screens

Each expanded graph is a three-dimensional linear-elastic beam–column frame with rectangular members, fixed bases, $E = 30$ GPa, Poisson ratio 0.2, and concrete unit weight 25 kN/m^3. Slab pressure is assigned to beams; plate bending, rigid-diaphragm behavior, and seismic response are outside the model. Story height is 3 m and slab thickness is 0.12 m. Under the complete-floor loading rule, each floor is one solid polygon: slab dead pressure is 3 kPa, additional dead pressure is 2 kPa, and live pressure is 2 kPa. The full floor area carries these imposed pressures across wall rings and column locations, counts, and footprints. The pressure resultant is held fixed as the beam graph changes. Beam-line tributary areas come from deterministic Voronoi cells of centerline samples no farther than 1.5 m apart, clipped to the floor and balanced to its full area. Member self-weight changes with section size; beam–slab and column–slab intersections are counted once. At each story, signed 0.5 kPa wind pressure acts on the floor bounding-box width perpendicular to each global horizontal direction times the 3 m story height. The resulting story force is divided equally among column-head joints at that level. Its total is fixed for a project, while its nodal allocation follows the column layout. Thus column edits can change the wind-force application centroid and torsional moment despite the fixed force resultant. Across the regional endpoints, the application centroid shifts 0.453–3.487 m from its parent; for U-plan, the $+X$ per-story torsional moment changes by $+23.75$ kN m. The supplement gives all 24 parent and endpoint centroids and moment arms. The independent OpenSees solve uses the same prescribed load vector.

For this template and a section schedule s, equilibrium for load case ℓ is

$$K(T, s)u_\ell = f_\ell(T, s), \qquad \ell \in \mathcal{L}. \tag{4}$$

Here K is the assembled elastic stiffness, u_ℓ the displacement vector, and f_ℓ the applied load vector for load case ℓ. The five SLS cases are $G + Q$, $G + Q \pm W_x$, and $G + Q \pm W_y$. The five ULS cases use the same signs and directions with coefficients $1.35G + 1.5Q \pm 1.5W_x$ or $1.35G + 1.5Q \pm 1.5W_y$, plus $1.35G + 1.5Q$. A CPU sparse factorization solves the ten right-hand sides together. Its custom implicit vector–Jacobian product solves the transposed free-degree system and propagates sensitivities through element stiffness and applied-load assembly.

For load case ℓ, let $w_{e\ell}$ be the maximum absolute vertical beam deflection across the 21 evaluation stations, and let $\theta_{c\ell}$ be the maximum absolute horizontal end-translation difference divided by column-segment length. For ULS, $q \in \{i, j\}$ indexes a column segment's two ends; $N^+_{cq\ell}$ is its compression demand, $M_{y,cq\ell}$ and $M_{z,cq\ell}$ are absolute end-moment demands, and $N_{t,cq\ell}$ is its

tensile axial demand. The column-end interaction is $U_{N,cq\ell} = N^{+}_{cq\ell}/N_{Rd,c} + |M_{y,cq\ell}|/M_{Rd,y,c} + |M_{z,cq\ell}|/M_{Rd,z,c}$. Hard acceptance uses the following normalized response screens in every load case:

$$\max_{e} w_{e\ell}/0.020\,\mathrm{m} \le 1, \qquad \text{SLS beam vertical envelope,} \tag{5}$$

$$\max_{c} \theta_{c\ell}/(1/250) \le 1, \qquad \text{SLS column-segment drift,} \tag{6}$$

$$U_{M,e\ell} = U_{M_y,e\ell} + U_{M_z,e\ell} \le 1, \qquad \text{ULS beam flexure,} \tag{7}$$

$$U_{V,e\ell} = U_{V_y,e\ell} + U_{V_z,e\ell} \le 1, \qquad \text{ULS beam shear,} \tag{8}$$

$$\max_{c,q} U_{N,cq\ell} \le 1, \qquad \text{ULS column interaction,} \tag{9}$$

$$\max_{c,q} N_{t,cq\ell} \le 10^{-6}\,\mathrm{N}, \qquad \text{ULS column tension screen.} \tag{10}$$

The drift maximum includes the first segment above the fixed base. In the finite-sizing expression below, $g_\ell(T, s)$ denotes these screen margins (normalized demand minus its limit); all entries must be nonpositive in every load case. Beam flexural and shear ratios add two directional proxy ratios. For $r \in \{y, z\}$, $(b_y, h_y) = (b, h)$ and $(b_z, h_z) = (h, b)$; then $d_r = h_r - c_{cov} - \phi_{st} - \phi_l/2$, $A_{s,r} = \rho_{max} b_r d_r$, $a_r = f_y A_{s,r}/(\alpha_1 f_c b_r)$, $M_{R,r} = f_y A_{s,r}(d_r - a_r/2)$, and $U_{M_r,e\ell} = M_{d,r,e\ell}/M_{R,r,e}$. Here $\rho_{max} = \min(0.025, \xi\alpha_1 f_c/f_y)$ and $\xi = 0.5$. The directional shear proxy is $V_{c,r} = 0.7 f_t b_r d_r$ and $U_{V_r,e\ell} = \max\{\max(V_{d,r,e\ell} - V_{c,r}, 0)/(f_{yv} d_r), A_{v,min}/s_{min}\}/(A_v/s)_{max}$. The minimum stirrups use $s_{min} = 0.20\,\mathrm{m}$, and the configured maximum is $(A_v/s)_{max} = 0.004\,\mathrm{m}$. The detailing inputs are 35 mm cover, 8 mm stirrups, assumed 20 mm longitudinal bars, and two-legged 8 mm minimum stirrups at 200 mm.

For square columns $A = bh$, $A_s = 0.04A$, and

$$N_{Rd} = 0.8 f_c (A - A_s) + f_y A_s, \tag{11}$$

$$M_{Rd,y} = 0.15 f_c b h^2 + 0.5 f_y A_s h, \tag{12}$$

$$M_{Rd,z} = 0.15 f_c h b^2 + 0.5 f_y A_s b. \tag{13}$$

The provisional strength inputs are $f_c = 14.3\,\mathrm{MPa}$, $f_t = 1.43\,\mathrm{MPa}$, $f_y = f_{yv} = 360\,\mathrm{MPa}$, and $\alpha_1 = 1$. These formulas serve as implementation screens rather than code design checks.

The independent OpenSees comparison covers all translational degrees of freedom, support-reaction components, and twelve local end-force components per member in each load combination. Agreement requires a maximum absolute translation difference below 10^{-8} m plus 10^{-5} times the block's largest reference translation magnitude, with a 10^{-12} m scale floor. For reactions and local forces, the corresponding rule is 10^{-3} in each output unit plus 10^{-5} times the block's largest magnitude, with a unit scale floor. The response screens are then recomputed from the OpenSees quantities using the same declared formulas. Across the 60 initial project–load-case comparisons, maximum absolute discrepancies are 4.16×10^{-14} m in translation and 4.92×10^{-6} in the reaction or local-force output units (N or N m). The highest independent screen ratio is 0.99868.

A.3 Catalog sizing and finite evaluation

The beam catalog (width, depth; metres) is

$$\mathcal{C}_b = \{(0.20, 0.40), (0.20, 0.50), (0.25, 0.60), (0.30, 0.70), (0.35, 0.80), (0.40, 1.00)\}.$$

Square-column side lengths are $(0.40, 0.50, 0.60, 0.70, 0.80, 1.00, 1.20)$ m. Each column's allowed entries are filtered by its complete footprint. The reported frame quantity is gross member volume, where b_e, h_e, and L_e are member width, depth, and length:

$$V(T, s) = \sum_{e \in E_{3D}(T)} b_e h_e L_e. \tag{14}$$

It omits slab, reinforcement, connections, foundations, and cost. The initial hard state uses the declared baseline sections (0.25×0.60 m beams and 0.50×0.50 m columns), projected only where

a column footprint fails. Feasible descent starts at the largest admissible catalog entries. Repeated floor beams share one decision, and each continuous column is grouped in three-story bands.

At each state the differentiable risk is the sum, over load cases and screen families, of the mean fourth powers of member-level dimensionless ratios: here e ranges over beams, c over column segments, and (c, q) over the two column-end checks per segment.

$$\begin{aligned} R = & \sum_{\ell \in \mathcal{L}_{SLS}} \left[\mathrm{mean}_e (w_{e\ell}/0.020)^4 + \mathrm{mean}_c (250\theta_{c\ell})^4 \right] \\ & + \sum_{\ell \in \mathcal{L}_{ULS}} \left[\mathrm{mean}_e \, U_{M,e\ell}^4 + \mathrm{mean}_e \, U_{V,e\ell}^4 \right. \\ & \left. + \mathrm{mean}_{c,q} \, U_{N,cq\ell}^4 \right]. \end{aligned} \tag{15}$$

The one-entry group rank score is $q_j = \Delta V_j / [10^{-3} + \max(0, \nabla R(s)^\top \Delta s_j)]$, where ΔV_j is predicted material saving. The method proposes up to 12 high-ranked groups together. A rejected hard batch is halved; acceptance requires passing all ten load cases and lowering gross volume. We checked the selfweight-sensitive derivative on four retained U-plan and Radial states, including two edited states with maximum raw screen ratios above 0.9998. All 24 sampled directional derivatives agreed with centered finite differences at a 10^{-5} m step under an absolute tolerance of 5×10^{-5} plus 0.5% relative tolerance. Reassembling dimension-dependent gravity loads and solving all ten load cases reproduced the base risk values. The seven sampled one-entry catalog reductions failed exact hard assessment; their actual risk increases exceeded the local linear predictions. Every risk-gradient factorization and hard trial consumes the budget. The finite result is

$$\widehat{V}_B(T) = \min_{s \in \mathcal{S}_B(T):\, g_\ell(T,s) \le 0 \ \forall \ell} V(T, s), \tag{16}$$

where $\mathcal{S}_B(T)$ is the set of schedules actually assessed. The result is a finite-search minimum over this set. At most 81 sizing factorizations and 600 s are allowed. If the largest admissible start fails, the finite search records that no feasible schedule was found. A complete topology evaluation also charges one initial mechanics factorization and one feedback factorization, each solving ten right-hand sides.

A.4 Action grammar and feedback

An edit query returns at most four ordered typed actions: move a joint, toggle its column status, add/remove a beam, split a beam with a new joint, or remove a joint with an admissible reconnection. The batch is applied atomically and revalidated for incidence, crossings, spatial fit, column-connected beam components, and the 50 mm coordinate grid. A trajectory may increase either the maximum or median 1 m floor-sample distance to beams by at most 0.05 m from its initial parent. Query records distinguish failed policy calls, invalid actions, coverage failures, and evaluated proposals. Failed requests and rejected proposals consume their query slot. After each accepted proposal the same full-floor evaluator performs a fresh sizing run. Incumbent ranking prefers feasible designs and orders them by volume. Infeasible designs are ordered by maximum violation, with volume breaking ties; failed assessments rank last.

The source-context packet provides the wall graph, declared surface, detected source regions and interfaces, a 50 mm action grid, section catalogs, and coordinate frames. Local axes are derived from wall-direction clusters modulo 90 degrees, with the surface centroid as origin; the canonical global frame is always retained. Global-only comparisons receive the global frame. Regional feedback reports connected member neighborhoods, governing load cases and ratios, selected sections, adjacent editable graph elements, admissible catalog choices, and remaining section growth. The witness view emphasizes the scalar objective, global checks, and worst-member witnesses. Both views receive the same architectural context and action grammar; no room-use or column-clearance semantics are inferred from wall labels.

A.5 U-PLAN COLUMN PLACEMENT AND SECTION FOOTPRINT

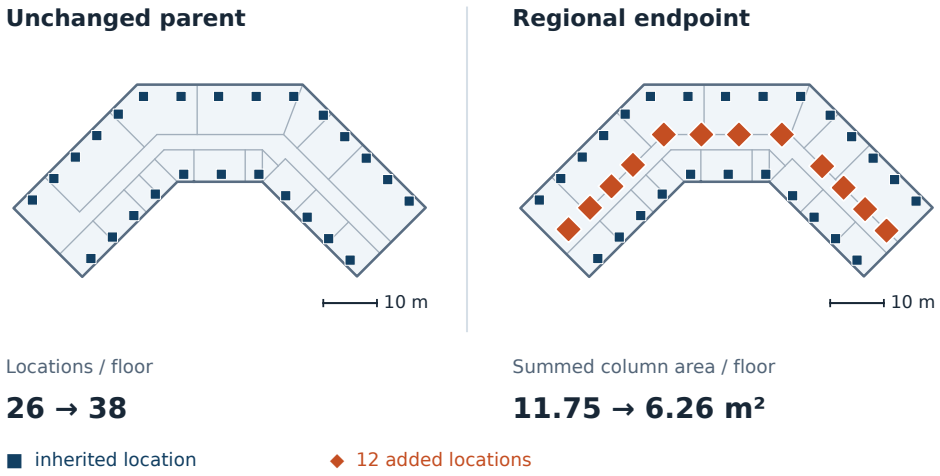

Figure 6. U-plan column locations in the unchanged parent and regional endpoint over first-storey source walls and the declared complete-floor outline. Navy squares mark inherited locations; rust diamonds mark twelve additions. Both plans use the same metric scale. Fixed-size symbols locate column centers; the values beneath compare summed selected column section area per floor after primary sizing (81 factorizations). Each selected location creates a continuous column through all three storeys.

The regional endpoint retains all 26 parent column locations and adds twelve. Eleven added centers are within 0.10 m of a source wall at each storey; the remaining added center is 3.022 m from its nearest source wall. All 38 column centers lie inside the selected source-derived floor outline. The selected sections reduce the summed cross-sectional area from 11.75 to 6.26 m^2 per floor (0.803% to 0.428% of the declared floor area), despite 36 additional vertical column segments across three stories.

A.6 ALL ASSESSED INITIAL DESIGNS

Table 5: All 37 evaluated framing-layout candidates. Col./floor counts column locations per typical floor; mean c.–wall is mean column-to-wall distance; Beam P95 is the 95th-percentile floor-point-to-beam distance. Recorded volume is shown for both passing and failing response screens.

Project	Family	V (m^3)	Feas.	Col./floor	Mean c.–wall (m)	Beam P95 (m)
Stepped	Initial model	785.86	Yes	29	0.83	2.95
Stepped	Rotated grid	666.27	Yes	27	1.50	2.85
Stepped	Rotated grid	810.90	Yes	33	1.83	2.55
Stepped	Rotated grid	510.21	Yes	16	1.77	3.95
Stepped	Wall aligned	1596.33	No†	34	0.00	12.35
Stepped	Perimeter seed	2520.69	Yes	79	0.20	4.80
Notched	Initial model	2533.36	Yes	44	0.59	2.60
Notched	Rotated grid	1802.52	Yes	36	0.98	3.30
Notched	Rotated grid	1347.00	Yes	28	0.56	2.80
Notched	Rotated grid	915.42	Yes	15	1.46	6.20
Notched	Wall aligned	4809.42	No†	62	0.00	2.65
Notched	Perimeter seed	4692.24	Yes	98	0.16	2.00
U-plan	Initial model	429.98	Yes	26	1.26	2.55
U-plan	Rotated grid	185.90	Yes	33	1.07	3.09
U-plan	Rotated grid	171.18	Yes	41	1.41	3.61

Continued on next page

Table 5 — continued from previous page

Project	Family	V (m^3)	Feas.	Col./floor	Mean c.–wall (m)	Beam P95 (m)
U-plan	Rotated grid	188.37	Yes	20	1.42	5.39
U-plan	Wall aligned	485.01	No†	27	0.01	6.64
U-plan	Perimeter seed	941.67	No†	66	0.19	3.64
Sloped-edge	Initial model	813.63	Yes	39	0.16	3.30
Sloped-edge	Rotated grid	833.18	Yes	46	1.32	3.43
Sloped-edge	Rotated grid	410.15	Yes	30	1.27	4.39
Sloped-edge	Rotated grid	424.95	Yes	21	1.27	5.06
Sloped-edge	Wall aligned	1482.28	No†	30	0.00	4.01
Sloped-edge	Perimeter seed	1128.41	Yes	64	0.21	3.15
Radial	Initial model	289.49	Yes	43	0.63	2.99
Radial	Rotated grid	339.12	Yes	58	1.80	2.80
Radial	Rotated grid	139.20	Yes	34	2.04	4.40
Radial	Rotated grid	168.66	Yes	30	1.36	4.05
Radial	Wall aligned	366.10	No†	22	0.00	7.92
Radial	Radial seed	124.21	Yes	24	1.30	12.23
Radial	Perimeter seed	608.45	No†	57	0.25	4.65
Winged	Initial model	1251.90	Yes	71	1.77	3.77
Winged	Rotated grid	1637.28	Yes	88	1.88	2.97
Winged	Rotated grid	922.41	Yes	57	1.82	3.79
Winged	Rotated grid	577.77	Yes	38	1.83	6.28
Winged	Wall aligned	2045.29	No†	65	0.00	6.88
Winged	Perimeter seed	3047.21	No†	132	0.18	6.12

Feas. refers to the declared geometry and linear-elastic frame checks. † marks a construction for which sizing found no passing start; its recorded volume is excluded from passing-design comparisons. Mean c.–wall is the mean distance from a column location to the nearest source-wall line. Beam P95 is the 95th-percentile distance from 1-m floor-grid points to the nearest beam centerline; it is not a slab span or capacity check.

A.7 Policy requests and numerical accounting

Table 6 summarizes all 60 primary model requests. Cached input is included in input-token totals. Rejected actions retain their query slots, while mechanics totals include only assessed proposals. All requests were configured for GPT-6 Sol/high.

Table 6. Model requests and assessed finite-element factorizations. The first four methods are paired decisions; the last two are three-query trajectories. Client seconds exclude structural evaluation.

Method	Requests	Input tokens	Cached input	Output tokens	Client s	Assessed	FE factorizations
Regional/local	6	243,479	32,256	12,402	268.9	4	332
Regional/global	6	241,781	32,256	8,710	207.5	5	415
Witness/local	6	138,517	32,256	12,609	274.4	5	415
Witness/global	6	136,811	32,256	10,986	240.3	4	332
Regional, sequential	18	734,143	137,088	28,512	668.8	17	1,331
Witness, sequential	18	421,010	120,960	35,938	785.0	16	1,248

The deterministic editor uses zero model requests. Its paired proposals use 338 recorded factorizations, and its eighteen sequential proposals use 854. Nine of all 84 primary query slots have no FE assessment record because their action or coverage check rejected the proposal; these slots are excluded from the FE sums and retained in the query denominators.

A.8 Deterministic comparison rules

The hotspot column-addition rule ranks noncolumn endpoints of the current material-hotspot beams, tries prefixes of four down to one in that order, and evaluates only the first graph- and coverage-admissible prefix. The five-family editor re-enumerates promotions, pruning, splits, exchanges, and

redundant beam removals after each accepted improvement. It assigns each admissible proposal the frozen score

$$q(a) = \frac{V_{\mathrm{beam},\downarrow} + V_{\mathrm{column},-} + V_{\mathrm{beam},-} - V^{\min}_{\mathrm{column},+} - V_{\mathrm{beam},\uparrow}}{\widehat{V}_{\mathrm{primary}}(T)}. \tag{17}$$

Here $\widehat{V}_{\mathrm{primary}}(T)$ is the parent volume under the primary 81-factorization sizing cap. The beam terms count each affected beam once and use the next catalog step down or up at its currently selected dimensions. Removed columns and beams contribute their selected gross volumes; an added column is charged at the smallest fitting cross-section along its full height. Ties use the highest current ratio of affected members, then a fixed family and action order. This proxy orders candidates before mechanics assessment. One admissible proposal is assessed per slot using the same 81-factorization sizing cap and feasible incumbent rule as the model trajectories.

Table 7. Persistent hotspot-promotion rule and post-hoc adaptive five-family editor.

Project	Hotspot (%)	Assessed	Adaptive (%)	Assessed
U-plan	47.33	3	47.33	3
Sloped-edge	0.00	2	1.43	3
Radial	0.00	3	0.44	3
Winged	0.00	0	6.34	3
Stepped	0.00	0	0.00	3
Notched	0.00	0	10.16	3
Mean	7.89		10.95	

The persistent hotspot-promotion rule uses 664 assessed batched factorizations across eight evaluations; the adaptive five-family editor uses 1,494 across eighteen. Both reuse the same initial parent assessments and have three proposal slots per project. They use zero model requests and zero finite-element prescreening calls. The hotspot-promotion rule uses fewer actual calls when no admissible column addition exists; the table matches proposal opportunity and per-topology caps.

Table 8. Post-hoc adaptive five-family editor and regional-feedback endpoint after common extended sizing.

Project	Unchanged (m^3)	Regional (m^3)	Adaptive (m^3)	Adaptive reduction (%)
U-plan	426.09	214.34	214.34	49.70
Sloped-edge	546.70	540.46	540.25	1.18
Radial	243.87	247.51	233.71	4.17
Winged	490.69	585.66	556.20	−13.35
Stepped	552.32	642.28	552.32	0.00
Notched	973.77	1667.42	1084.95	−11.42

Each graph receives at most 321 sizing factorizations. All eleven distinct parent/endpoint states pass independent OpenSees agreement and the declared response screens at both the primary and longer sizing caps: 22 ten-load-case checks, or 220 load-case runs. The Stepped endpoint is the unchanged parent.

Table 9. Column locations and selected cross-sectional area after common 321-factorization sizing. Regional denotes regional feedback; Five-family denotes the post-hoc adaptive five-family editor.

Project	Regional columns	Five-family columns	Regional area	Five-family area
U-plan	38	38	6.170	6.170
Sloped-edge	37	39	7.375	7.685
Radial	40	43	7.470	7.840
Winged	62	63	11.623	11.488
Stepped	25	29	6.120	6.153
Notched	40	39	19.221	9.933

Area is the mean across floors of the summed selected column cross-sectional area in m^2 per floor. Both graphs in each project retain the same loaded floor area and imposed pressure as their selected column sections change.

Table 10. Member-volume–column-location exchange rates at the common 321-factorization cap.

Project	$V_{\text{regional}} - V_{\text{adaptive}}$ (m^3)	$N_{\text{adaptive}} - N_{\text{regional}}$	λ^*_{col} (m^3/location)
Sloped-edge	0.204	2	0.102
Radial	13.795	3	4.598
Winged	29.457	1	29.457
Stepped	89.960	4	22.490

For the illustrative scalar objective $J = V + \lambda_{\text{col}} N_{\text{col}}$, the regional design has lower J in each row when $\lambda_{\text{col}} > \lambda^*_{\text{col}}$. Here V is whole-building gross beam-and-column volume and N_{col} is the number of column locations on the typical floor; the exchange rate is a volume equivalent across the modeled building, not a monetary price. U-plan is tied in both quantities, while the adaptive five-family editor has lower volume and fewer column locations on Notched.

Table 11. Common-budget member volume and column locations for all tested graph endpoints. Parent is unchanged; Witness denotes governing-witness feedback; Deterministic is the prespecified source-aware editor; Regional denotes regional feedback; Five-family is the post-hoc adaptive five-family editor.

Project	Parent	Witness	Deterministic	Regional	Five-family
U-plan	426.09/26	368.53/28	426.09/26	214.34/38	214.34/38
Sloped-edge	546.70/39	550.57/37	709.63/37	540.46/37	540.25/39
Radial	243.87/43	251.63/41	260.92/39	247.51/40	233.71/43
Winged	490.69/71	649.24/59	662.24/59	585.66/62	556.20/63
Stepped	552.32/29	538.76/27	727.46/26	642.28/25	552.32/29
Notched	973.77/44	1055.74/38	1425.31/40	1667.42/40	1084.95/39

Each entry lists whole-building gross beam-and-column volume (m^3) followed by the number of column locations on the typical floor, after sizing each fixed graph under the same 321-factorization cap. The regional endpoint is Pareto-nondominated among these five tested method states in every project except Notched. On U-plan it exactly matches the five-family graph. Equality alone does not imply nondomination if another design improves both measures.

A.9 MATCHED NUMERICAL-CALL CONTINUATION

Table 12. Unchanged-graph sizing at matched total factorization counts.

Project	Method	FE factorizations	Primary V (m^3)	Unchanged V (m^3)
U-plan	Deterministic	92	429.98	429.98
U-plan	Regional	332	226.46	426.09
U-plan	Witness	252	382.05	426.43
Sloped-edge	Deterministic	172	805.95	551.62
Sloped-edge	Regional	332	769.74	546.04
Sloped-edge	Witness	249	751.17	549.67
Radial	Deterministic	332	278.76	243.87
Radial	Regional	332	268.97	243.87
Radial	Witness	332	270.41	243.87
Winged	Deterministic	332	1169.82	490.69
Winged	Regional	332	1166.21	490.69
Winged	Witness	332	1144.37	490.69
Stepped	Deterministic	252	785.84	553.94
Stepped	Regional	332	698.61	552.32
Stepped	Witness	332	700.52	552.32
Notched	Deterministic	172	2415.58	1454.79
Notched	Regional	169	2382.65	1472.20
Notched	Witness	249	2218.21	1002.20

Note. FE factorizations comprise the 83-factorization initial evaluation and the actual factorizations for assessed proposals, up to 332 total. Deterministic denotes the prespecified source-aware editor. Independent OpenSees validation is reported separately. The unchanged continuation spends the same total allowance on the original graph. Its volume is read from the retained trace at that allowance, carrying forward a terminal value after all feasible one-entry reductions are exhausted. Primary volume is the selected endpoint after at most 81 sizing factorizations per topology.

A.10 EXTENDED FROZEN-GRAPH SIZING

All common 321-factorization reassessments reach their cap; the extended endpoint statuses below distinguish exhausted search from no remaining feasible one-entry group reduction.

Table 13. Frozen-graph sizing endpoints after at most 4,097 factorizations (gross member volume, m^3).

Project	Unchanged	Original deterministic	Regional	Witness
U-plan	417.88	417.88	209.67	363.16
Sloped-edge	523.16	632.44†	519.76	532.40
Radial	238.98	254.02	245.20	245.13
Winged	487.35	641.61	578.21	621.98
Stepped	546.63	707.32	615.73	531.56
Notched	916.65	1200.10†	1458.10†	1000.60†

Note. † marks an endpoint that reached the factorization cap before exhausting available reductions. The other 19 physical endpoints stopped after assessing every still-reducible group and finding no feasible one-entry reduction. Method-versus-same-project-unchanged volume ordering agrees between 321 and 4,097 factorizations in 18/18 pairs and between 81 and 4,097 factorizations in 5/18 pairs.

A.11 SEPARATE POLICY SAMPLES ON SLOPED-EDGE, WINGED, AND STEPPED

After the primary analysis, we repeated both three-query model policies from the same generated parents on Sloped-edge, Winged, and Stepped. The first-query prompts match the saved primary prompts exactly; the same architecture, geometry, 81-factorization sizing cap, and ten load cases were used. The requested model and effort were GPT-6 Sol/high. This run used Codex CLI 0.157.1 rather than the primary run's recorded 0.158.0-alpha.2, and the served model identity is unavailable, so we report the samples separately.

Table 14. Separate three-query samples from common parents.

Project	Parent (m^3)	Regional (m^3, %)	Witness (m^3, %)
Sloped-edge	813.63	758.41 (6.79%)	715.37 (12.08%)
Winged	1251.90	1186.51 (5.22%)	1155.65 (7.69%)
Stepped	785.86	716.77 (8.79%)	709.74 (9.69%)

All 18 model requests parsed; 16 proposals reached assessment, one action was rejected, and one failed the coverage guard. All final schedules pass the declared frame response screen. Across ten load cases per endpoint, independent OpenSees analyses agree with the primary solver and pass the declared screen in all 60 checks. Witness feedback gives lower member volume in all three new pairs; Stepped reverses the primary trajectory's regional–witness ordering. The supplement provides the separate machine-readable result record.