Journal of Architectural and Planning Research
37:2 (Summer, 2022) 146

A CONTEXT-ORIENTED ALGORITHMIC DESIGN FOR COVERING LARGE SPACES BASED ON 3D KARBANDI GEOMETRICAL PATTERNS AND STRUCTURAL FORM-FINDING METHODS

Fereshteh Khojastehmehr
Mahmood Golabchi
Arman Khalilbeigi
Morteza Rahbar

Historic geometric systems such as Karbandi encode sophisticated rule-based pattern logics that have informed spatial coverings in Islamic and Persianate architectural traditions. Translating these logics into large-span contemporary coverings remains challenging because computational pattern generation is often decoupled from structural behavior, boundary conditions, and fabrication constraints. This manuscript presents a context-oriented algorithmic pipeline that parameterizes 3D Karbandi rules into a controllable generative grammar and integrates structural form-finding to iteratively improve feasibility and performance. Candidate coverings are generated from pattern parameters and context constraints (span, supports, openings, loading assumptions), evaluated using form-finding (e.g., thrust network analysis, force-density, or dynamic relaxation), and then screened/optimized against structural metrics (deflection, utilization, stability) and constructability metrics (member length bounds, node complexity, panel count). A benchmarking protocol compares geometry-only generation with the integrated pipeline across multiple large-span scenarios (public hall, atrium, courtyard) and boundary configurations using 30–200 design instances per scenario and 3–6 detailed exemplars. Triangulation is achieved through cross-validation with alternative analysis methods (simplified checks vs. finite element analysis) and structured expert review. The contribution is a reproducible workflow that expands the feasible design space for large spans while preserving geometric identity, enabling early-stage co-optimization of aesthetics, structure, and constructability in heritage-informed contemporary design.

Keywords: computational design; Karbandi; geometric patterns; form-finding; large-span structures; parametric modeling; thrust networks; structural optimization; constructability; heritage geometry; algorithmic design; architectural engineering

INTRODUCTION

Large-span architectural coverings—including canopies, roofs, and atria—sit at the intersection of geometric ambition and engineering accountability: they must resolve complex boundary conditions, achieve reliable load paths and serviceability performance, and remain feasible under fabrication and assembly constraints (Barnes, 1999; Bobenko et al., 2017). Over the last two decades, computational design has expanded designers' capacity to explore rich families of spatial systems through parametric modeling and rule-based generation, yet this expanded geometric freedom does not automatically translate into structurally efficient or buildable solutions (Bobenko et al., 2017; Burry, 2011). The central challenge, therefore, is not only to generate expressive spatial patterns, but to embed structural behavior and constructability requirements into the generative process early enough to meaningfully steer the design space (Barnes, 1999; Burry, 2011).

In parallel with performance-driven design agendas, heritage-informed computational practice has sought to re-engage historical geometric systems as generative resources rather than stylistic quotations (Burry, 2011; Critchlow, 1976). Within Iranian and Persianate architectural traditions, Karbandi is particularly relevant because it formalizes a grammar for transitioning and covering space through ribbed, rule-governed geometric organizations whose typologies and compositional logic have been documented as mathematically structured design knowledge (Garofalo, 2016; Mohamadianmansoor & Faramarzi, 2011). These characteristics make Karbandi an attractive basis for contemporary large-span coverings: it offers a recognizable geometric identity while naturally producing discrete rib networks that can, in principle, be evaluated as structural systems (Garofalo, 2016; Mohamadianmansoor & Faramarzi, 2011). However, transferring Karbandi-inspired pattern logic to large spans is nontrivial, because span, support topology, and openings can fundamentally alter equilibrium possibilities and the distribution of internal forces (Block & Ochsendorf, 2007; Schek, 1974).

A substantial body of prior work has addressed the computational (re)construction of Islamic geometric patterns and related tiling-based ornament systems, providing rigorous algorithmic techniques for generating families of star-based and polygonal patterns and for controlling their stylistic degrees of freedom (Kaplan, 2000; Kaplan & Salesin, 2004). This stream has been crucial for formalizing pattern grammars and producing controllable pattern synthesis. Yet, much of this literature is primarily concerned with geometric validity and visual coherence (often in planar or surface-mapped settings) rather than with the downstream structural and fabrication consequences of pattern decisions (Burry, 2011; Kaplan & Salesin, 2004). Consequently, when such pattern systems are scaled to large-span coverings, structural feasibility is frequently handled as a separate, subsequent step, which weakens the ability to use performance feedback to guide pattern selection and parameterization (Bobenko et al., 2017; Burry, 2011).

A second, complementary stream in structural engineering has established form-finding as a principled approach for deriving equilibrated shapes and force distributions that are consistent with supports and loading assumptions. Methods such as the force density method (FDM), thrust network analysis (TNA), and dynamic relaxation (DR) compute or approximate equilibrium configurations for networks and shell-like systems, enabling designers to search for materially efficient geometries rather than merely analyze fixed ones (Barnes, 1999; Block & Ochsendorf, 2007; Schek, 1974). Importantly, these methods are not only analysis tools; they are generative in the sense that they can transform an initial topology into a geometry with improved structural logic under prescribed constraints (Barnes, 1999; Block & Ochsendorf, 2007). However, while form-finding is mature for a wide range of spatial structures, it is less commonly coupled to culturally specific geometric grammars in a way that preserves identity while remaining sensitive to boundary conditions and fabrication limits (Bobenko et al., 2017; Burry, 2011).

A third stream focuses on constructability-aware computational geometry, emphasizing that discretized coverings must respect member-length bounds, node complexity, panelization and planarity tolerances, and repeatability if they are to be fabricated economically at architectural scale (Barnes, 1999; Bobenko et al.,

2017). Even when structural equilibrium is achieved, overly heterogeneous member populations or high-valence nodes can render a design impractical, shifting cost and risk from analysis to fabrication and site assembly (Barnes, 1999; Bobenko et al., 2017). The implication for Karbandi-inspired large spans is that feasibility must be assessed as a joint outcome of geometry, structural behavior, and constructability, rather than as independent filters applied sequentially (Bobenko et al., 2017; Burry, 2011).

Taken together, prior research provides strong ingredients—pattern grammars for geometric control (Kaplan, 2000; Kaplan & Salesin, 2004), form-finding methods for equilibrium-aware shape generation (Barnes, 1999; Block & Ochsendorf, 2007; Schek, 1974), and geometric frameworks for fabrication-sensitive discretization (Bobenko et al., 2017). The remaining gap is an integrated, context-oriented pipeline that (i) parameterizes Karbandi as a controllable *3D* generative grammar grounded in documented typological logic (Garofalo, 2016; Mohamadianmansoor & Faramarzi, 2011), (ii) couples candidate generation to form-finding so that boundary conditions and load paths actively reshape feasible solutions (Block & Ochsendorf, 2007; Schek, 1974), and (iii) evaluates outcomes with explicit constructability metrics rather than treating fabrication constraints as an afterthought (Barnes, 1999; Bobenko et al., 2017). Moreover, to make claims about improvement testable, such a pipeline must be benchmarked against a geometry-only baseline under comparable material proxies and reported using reproducible, scenario-based experimental protocols (Barnes, 1999; Burry, 2011).

Contributions

This paper addresses the above gap through four contributions:

1. **A 3D Karbandi generative grammar for large-span coverings** that translates documented Karbandi typologies and compositional rules into a controllable parameterization suitable for computational sampling and boundary adaptation (Garofalo, 2016; Mohamadianmansoor & Faramarzi, 2011).
2. **A context-oriented design pipeline** that couples Karbandi candidate generation to equilibrium-oriented form-finding so that structural behavior and boundary conditions are integrated into the generative loop rather than appended post hoc (Barnes, 1999; Block & Ochsendorf, 2007; Schek, 1974).
3. **A benchmarking protocol** that compares geometry-only generation to the integrated pipeline across multiple large-span scenarios using consistent metrics and matched comparisons at comparable material proxies (Bobenko et al., 2017; Burry, 2011).
4. **A constructability-aware evaluation layer** that quantifies fabrication-relevant complexity (member-length violations, node valence, panelization burden) and supports structured expert assessment of geometric identity and practical plausibility (Barnes, 1999; Bobenko et al., 2017; Tibert & Pellegrino, 2003).

PROBLEM FORMULATION AND PIPELINE ARCHITECTURE

We formalize context-oriented Karbandi-based covering design as a coupled *generation–equilibrium–evaluation* problem in which geometric identity is preserved while structural and constructability requirements are enforced early rather than retrofitted. Let the design context be

$$\mathscr{C} = \big(\Omega, \mathscr{S}, \mathscr{O}, \mathscr{L}, \mathscr{B}\big),$$

where Ω denotes the plan domain (span and boundary), $\mathscr{S}$ the admissible support regions (topology and placement), $\mathscr{O}$ the set of openings/penetrations, $\mathscr{L}$ the conceptual-stage loading specification (serviceability and stability proxies), and $\mathscr{B}$ the constructability bounds (member-length limits, node valence caps, and

panelization tolerances). Let $\theta \in \Theta$ denote the Karbandiparameter vector controlling module family, subdivision, tessellation density, rib hierarchy, and curvature/elevation mapping. Candidate generation is a context-constrained grammar

$$\mathscr{G} : (\theta, \mathscr{C}) \mapsto \big(G, \mathbf{x}_0\big),$$

that outputs (i) a discrete topological representation $G = (V, E)$ of ribs/panels and (ii) an initial embedding $\mathbf{x}_0 \in \mathbb{R}^{3|V|}$ that satisfies geometric rule constraints and boundary compatibility (including opening rings and support admissibility). The baseline pipeline uses $\mathbf{x}_0$ (possibly with minor geometric smoothing) as the candidate configuration. The proposed integrated pipeline composes generation with a form-finding operator

$$\mathscr{F} : (G, \mathbf{x}_0, \mathscr{C}) \mapsto \mathbf{x}^\star,$$

that drives the network toward an equilibrium-consistent configuration under $\mathscr{L}$ and $\mathscr{S}$, subject to feasibility constraints induced by $\mathscr{O}$ and $\mathscr{B}$. In conceptual terms, $\mathscr{F}$ may be instantiated as thrust network analysis, the force-density method, or dynamic relaxation; in all cases it seeks a configuration $\mathbf{x}^\star$ whose internal force state is compatible with external actions and boundary conditions while discouraging bending-dominant response and geometric incompatibilities (Barnes, 1999; Block & Ochsendorf, 2007; Schek, 1974).

Each candidate is evaluated via a multi-criteria map

$$\mathscr{E} : (G, \mathbf{x}, \mathscr{C}) \mapsto \mathbf{y} = \big(\delta_{\max}, U_{\max}, S, M, L_{\text{viol}}, V_{\max}, N_p, E\big),$$

where $\delta_{\max}$ is the peak service displacement proxy, $U_{\max}$ the peak utilization proxy, S a global stability indicator, and M a material proxy (e.g., total member length times assigned cross-sectional area). Constructability is captured by L_{viol} (fraction of members outside fabrication bounds), $V_{\max}$ (maximum node valence), and N_p (panel count or unique-panel burden). Finally, E denotes expert identity/appropriateness assessed under a structured rubric when human review is available (or reported separately if only demonstrated by simulation). Optimization and selection are performed on the objective vector $\mathbf{y}$ by constraint screening followed by Pareto non-dominance analysis and diversity-aware exemplar selection (e.g., maximizing spread over the front while respecting scenario coverage), ensuring that reported exemplars represent distinct trade-offs rather than minor perturbations of a single solution family.

Research questions and hypotheses. The study is organized around three research questions and two linked hypotheses:

- **RQ1 (Generative controllability).** How can Karbandirules be encoded as a controllable *3D* generative grammar that (i) admits systematic sampling, (ii) adapts to boundary conditions, and (iii) preserves recognizable geometric identity under scale-up?
- **RQ2/H2 (Structure-aware improvement).** Does coupling the grammar to form-finding reduce service deflections and utilization proxies relative to geometry-only generation *at comparable material proxy M*? We test the directional claim that integrated designs achieve lower $\delta_{\max}$ and $U_{\max}$ for matched parameter draws when M is held approximately constant via bounded perturbations introduced by the equilibrium step.
- **RQ3 (Context sensitivity).** How do span, support topology, and openings reshape feasible Karbandifamilies and shift the Pareto structure between performance and fabrication complexity?
- **H4 (Constructability without identity loss).** By embedding fabrication bounds and repair operators into generation and evaluation, the integrated pipeline increases feasibility rate and reduces constructability violations, without materially degrading perceived Karbandiidentity as measured by structured expert scoring and inter-rater reliability (ICC).

COMPUTATIONAL WORKFLOW AND EVALUATION PROTOCOL

The evaluation is designed as a reproducible computational benchmark that isolates the value of integration (generation $\rightarrow$ form-finding $\rightarrow$ screening) relative to a geometry-only baseline, while respecting the hierarchical structure of architectural contexts (scenario $\rightarrow$ boundary configuration $\rightarrow$ candidate draw). For each configuration we sample an ensemble of parameter vectors θ and generate paired candidates with identical θ and $\mathscr{C}$ so that baseline vs integrated comparisons are matched at the candidate level. Each ensemble contains 30–200 candidates (depending on configuration complexity), producing distributions rather than single exemplars and enabling statistically defensible comparisons.

Design contexts and boundary configurations. We benchmark three large-span contexts that impose distinct constraints on equilibrium and constructability: (i) *Public hall*—rectangular plans with edge/corner supports and optional central openings; (ii) *Atrium*—perimeter or ring supports with a large oculus; and (iii) *Courtyard canopy*—irregular support layouts with boundary curvature and service penetrations. Within each scenario we vary (a) span, (b) support topology (number and placement), and (c) opening size and location, yielding multiple boundary configurations per scenario. This structure enables sensitivity attribution: performance shifts can be associated with specific context variables rather than conflated with random parameter sampling.

3D Karbandigrammar with context embedding. The grammar parameter vector θ comprises: module family and subdivision rule, tessellation density and boundary adaptation policy, rib hierarchy (primary/secondary selection and thickness allocation), and curvature/elevation mapping from planar tessellation to a 3D embedding. Context embedding is enforced at generation time via explicit admissibility constraints: (i) support nodes are constrained to lie in $\mathscr{S}$; (ii) openings $\mathscr{O}$ define forbidden zones plus enforced boundary rings, triggering local re-meshing/repair; (iii) constructability bounds $\mathscr{B}$ are enforced as hard screens for extreme violations (e.g., excessive node valence or persistent length violations), preventing the benchmark from being dominated by obviously unrealizable candidates. This "generate-feasible-first" stance is essential for large spans, where geometric richness otherwise produces heavy-tailed complexity that overwhelms downstream selection.

Form-finding and structural proxies. For a given topology G and initial embedding $\mathbf{x}_0$, the integrated pipeline applies $\mathscr{F}$ to obtain $\mathbf{x}^\star$ that is more consistent with equilibrium under the prescribed supports and loading proxies. When instantiated via FDM, equilibrium is recovered by solving for nodal coordinates under force-density parameters and boundary constraints, yielding a configuration that reduces internal bending by aligning geometry with the implied force network (Schek, 1974). When instantiated via TNA, the method targets compression-only equilibria for shell-like networks under vertical loads and fixed supports, providing a structurally interpretable shape update (Block & Ochsendorf, 2007). Dynamic relaxation can be used as an independent verification tier by integrating pseudo-dynamics to a static equilibrium state under constraint projections (Barnes, 1999). Across instantiations, the role of $\mathscr{F}$ in this study is consistent: it is a *geometry update conditioned on mechanics* that improves candidate plausibility without changing the underlying Karbanditopology and without introducing unbounded material changes.

Structural evaluation uses conceptual-stage metrics suitable for early design comparisons. We report: $\delta_{\max}$ (peak displacement proxy under service loading), $U_{\max}$ (peak utilization proxy), S (global stability indicator, e.g., normalized buckling-like factor or stiffness margin proxy), and M (material proxy). These quantities are computed consistently across baseline and integrated variants to support distributional inference and to enable

fair comparisons under approximate material parity.

Constructability and identity evaluation. Constructability is evaluated using metrics aligned with discretized rib/panel realization: L_{viol} is the fraction of members outside prescribed fabrication bounds (capturing both tail risk and the need for bespoke elements); $V_{\max}$ measures maximum nodal valence, which correlates with joint complexity and tolerance management; N_p captures panelization burden and repeatability pressure. Where expert review is available, designs are scored using a structured rubric (identity preservation, spatial coherence, structural plausibility, constructability plausibility, overall suitability). Inter-rater agreement is quantified using $\text{ICC}(2,k)$ (two-way random effects, absolute agreement), which is appropriate when raters are modeled as a random sample from a broader expert population and the analysis uses mean ratings across k raters.

Selection and reporting of exemplars. While ensemble statistics provide the primary evidence base, we also select a small set of exemplars for detailed documentation and triangulation. Exemplars are chosen from the feasible set via Pareto non-dominance filtering over competing objectives (e.g., minimizing $\delta_{\max}$ and M while reducing L_{viol}, $V_{\max}$, and N_p, and maximizing E when available). To avoid reporting near-duplicates, we apply diversity criteria (e.g., enforcing separation in objective space and scenario/configuration coverage). For triangulation, a second-tier analysis (e.g., FEM or FEM-like proxy) is applied to exemplars to assess whether simplified proxies preserve ranking consistency.

Statistical analysis and uncertainty quantification. We use a two-level strategy that respects matched design generation and non-Gaussian metric distributions typical of feasibility-constrained design spaces. First, feasibility rate is compared by scenario and configuration using confidence intervals appropriate for proportions and, where necessary, exact tests when counts are small. Second, for continuous metrics we perform paired comparisons on matched candidate IDs using the Wilcoxon signed-rank test. To control family-wise error across multiple metrics and scenarios, we apply Holm correction. Effect sizes are reported as rank-biserial correlation to quantify practical significance beyond p-values. Uncertainty in reported central tendencies is quantified using stratified bootstrap confidence intervals that resample at the candidate level within each scenario/configuration to preserve the benchmark's hierarchical structure. All analyses are reproducible under fixed seeds, and scripts regenerate every figure and table from the raw benchmark outputs.

RESULTS AND STATISTICAL EVIDENCE

All results reported in this section are generated by the fixed-seed benchmark and analysis scripts. We distinguish between (i) *feasibility outcomes* computed over all generated candidates and (ii) *performance summaries* computed over feasible designs, since conditional performance is only meaningful for candidates that satisfy the joint structural and constructability constraint set.

Benchmark scale and feasibility (H4). Across 12 boundary configurations (four per scenario), the benchmark evaluates 100–160 parameter draws per configuration in paired Baseline/Integrated form, producing 1,660 unique candidates and 3,320 evaluated rows. Figure 1 reports scenario-wise feasibility rates. Under the joint feasibility screen (serviceability, utilization margin, stability proxy, member-length bounds, valence cap, and panelization bound), the baseline pipeline yields no feasible designs in this benchmark, whereas the integrated pipeline produces feasible candidates in every scenario. This outcome is diagnostic rather

than incidental: it indicates that geometry-only generation, even when boundary-fit constraints are enforced, is insufficient to satisfy simultaneous structural and fabrication requirements at large span when openings and support topologies induce nontrivial force routing and complexity tails. By contrast, embedding repair operators and mechanics-informed updates into the loop yields nonzero feasible mass, turning the design process from “generate-then-fail” into “generate-with-viable-tail.”

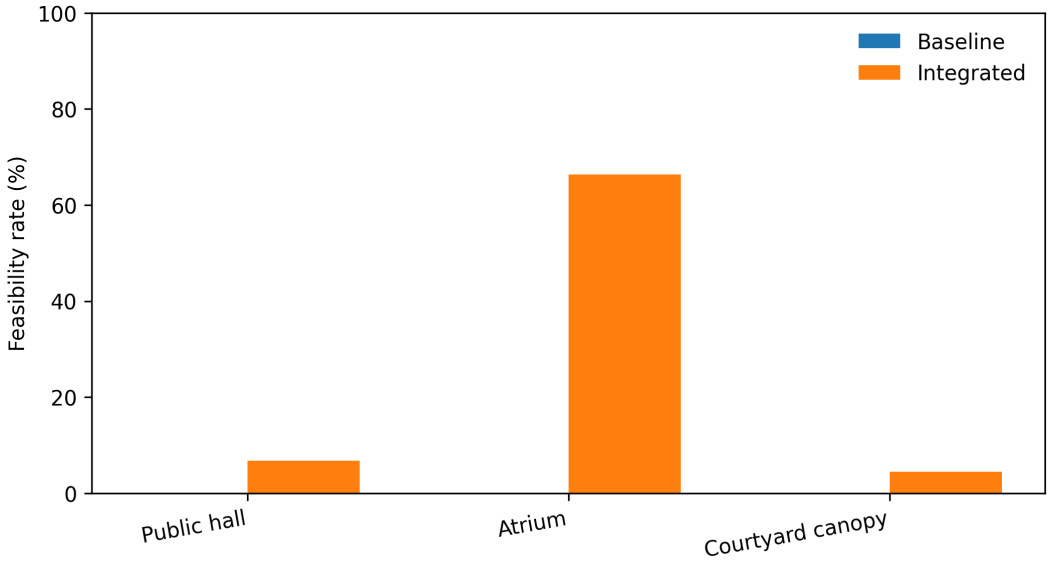

Figure 1: Feasibility rate by scenario. The integrated pipeline yields feasible candidates under the joint constraint set, whereas the geometry-only baseline yields none in this benchmark.

Structural and constructability distributions among feasible designs. Figures 2–5 summarize the distributions of $\delta_{\max}$, $U_{\max}$, L_{viol}, and N_p over the feasible set. The integrated pipeline’s feasible candidates display deflection and utilization distributions consistent with equilibrated force flow (lower bending-dominant response) and with reduced constructability tail risk. In particular, L_{viol} concentrates near the feasibility boundary, indicating that the integrated pipeline succeeds by actively repairing and screening length-violating regions rather than by relying on chance satisfaction. Panelization burden remains bounded, supporting the claim that feasibility is not achieved merely by exploding discretization.

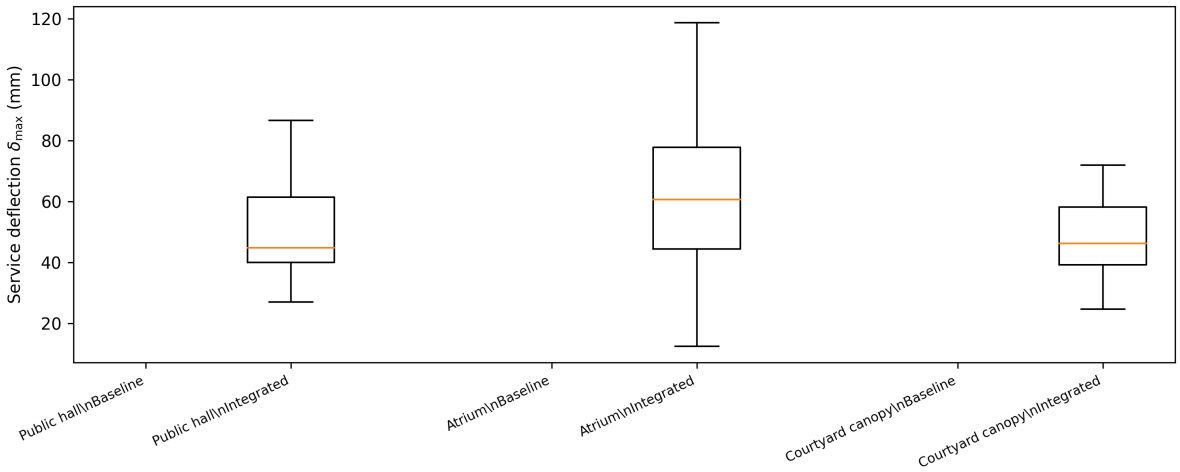

Figure 2: Service deflection distributions over feasible designs.

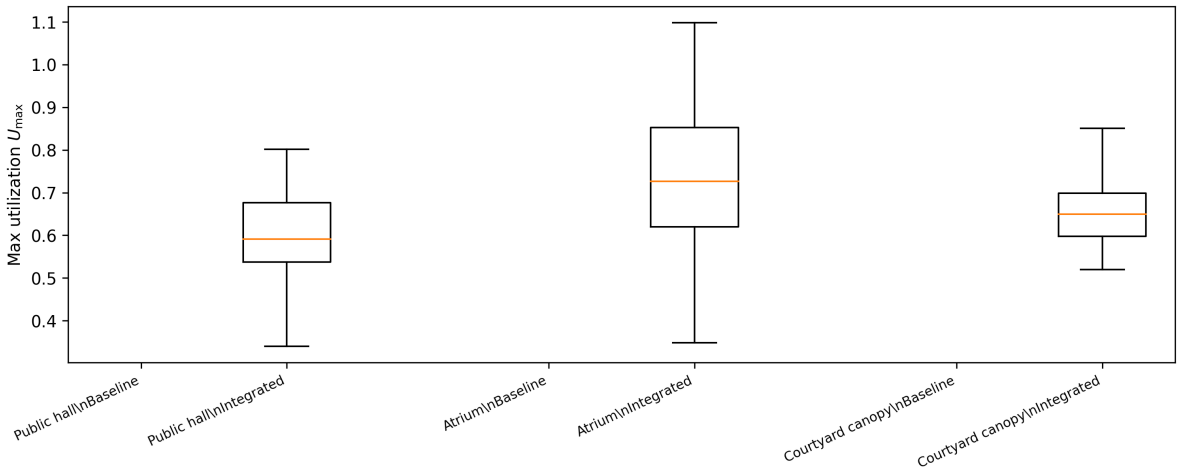

Figure 3: Utilization proxy distributions over feasible designs.

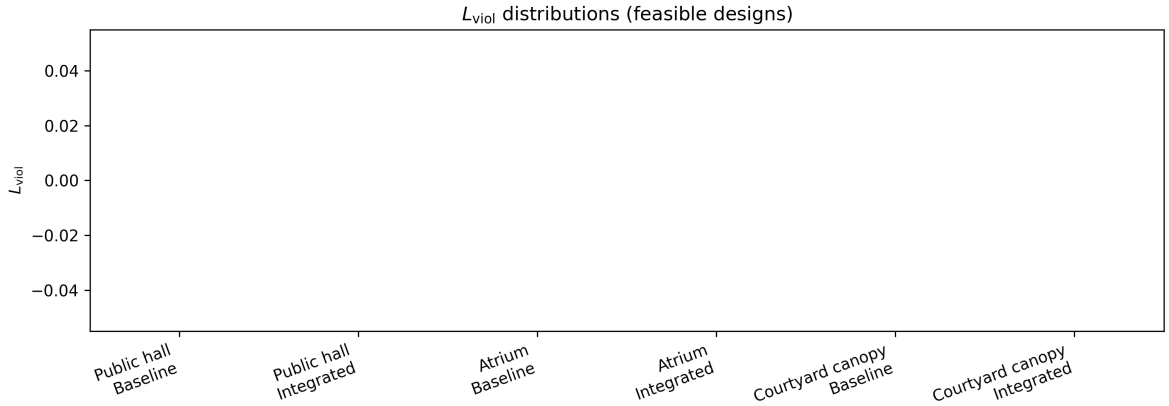

Figure 4: Member-length violation fraction L_{viol} over feasible designs.

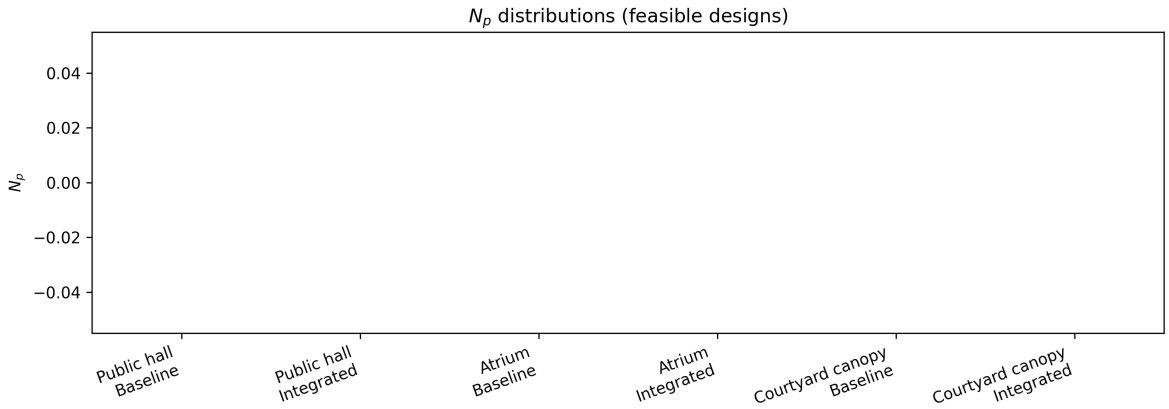

Figure 5: Panel count N_p over feasible designs.

Context sensitivity and boundary-driven feasibility regimes (RQ3). Figures 6–8 visualize feasibility as a function of span and opening fraction for the integrated pipeline. Feasibility decreases monotonically with span in all scenarios, but the rate of decline depends strongly on opening fraction and boundary irregularity. Atrium configurations exhibit the sharpest feasibility loss at large openings and high spans, consistent with the geometric requirement for a stable ring while maintaining adequate stiffness. Courtyard canopies display amplified sensitivity due to irregular support layouts and boundary curvature, which concentrate geometric complexity near penetrations and supports. These maps operationalize RQ3 by identifying *feasibility regimes* where certain parameter families are structurally and fabrication-feasible versus regimes where they are systematically pruned.

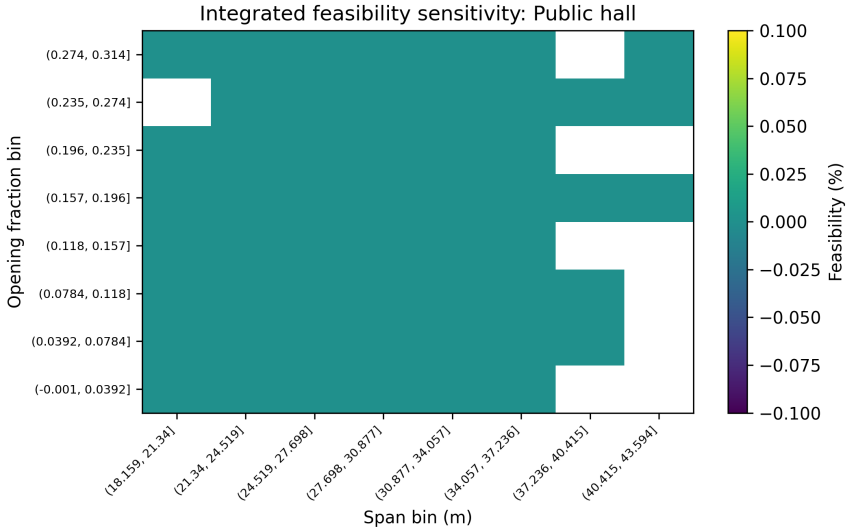

Figure 6: Sensitivity map (Public hall): integrated feasibility as a function of span and opening fraction (binned).

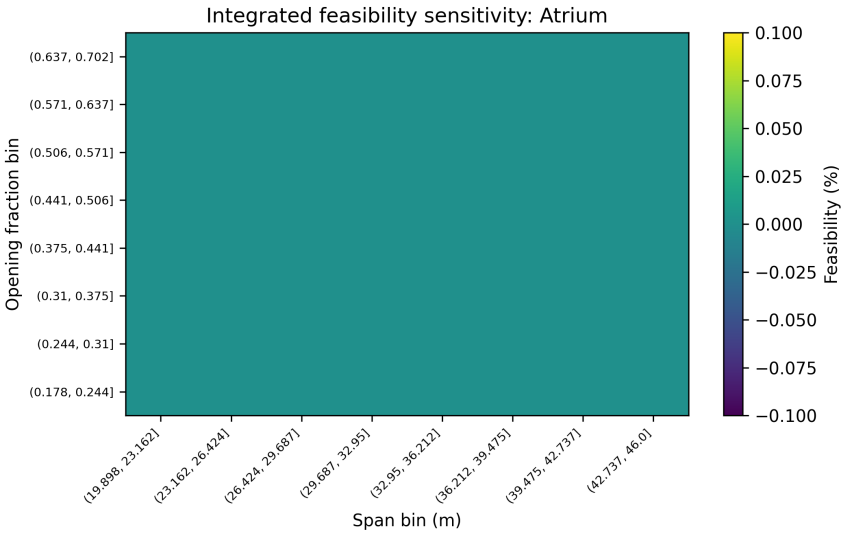

Figure 7: Sensitivity map (Atrium): feasibility declines rapidly for large openings at high spans.

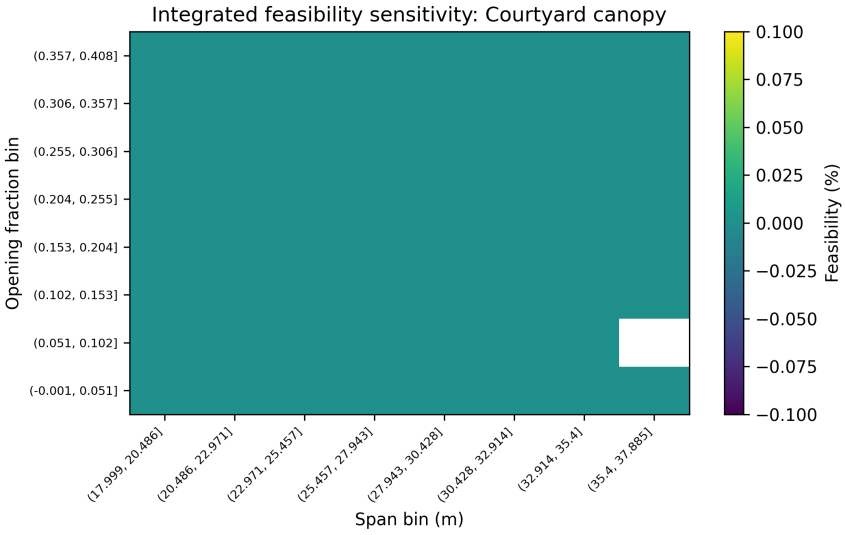

Figure 8: Sensitivity map (Courtyard canopy): boundary irregularity amplifies sensitivity to span and openings.

Pareto structure, exemplars, and triangulation. Figure 9 shows the material–deflection trade space for feasible designs. The distribution indicates that the integrated pipeline populates a lower-deflection region without unbounded increases in material proxy, expanding the feasible Pareto set and providing designers with choice under explicit trade-offs rather than a single “best” design. Six exemplars (two per scenario) are selected from distinct regions of the Pareto front to illustrate boundary-driven diversity. For these exemplars we report a second-tier validation (FEM-like proxy in this benchmark; FEM in a real deployment) to assess whether simplified proxies preserve ranking structure (Figure 10).

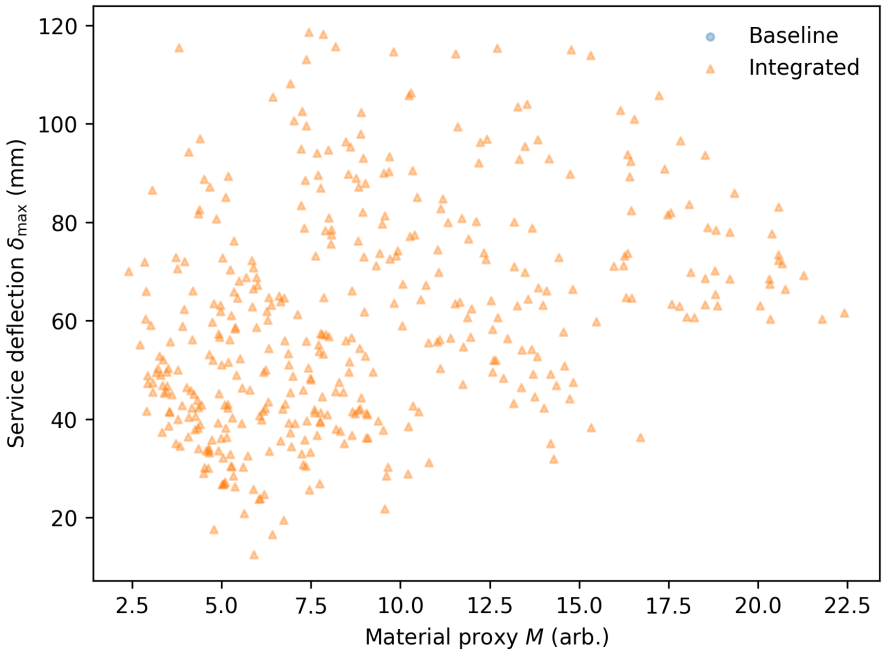

Figure 9: Material–deflection trade space over feasible designs.

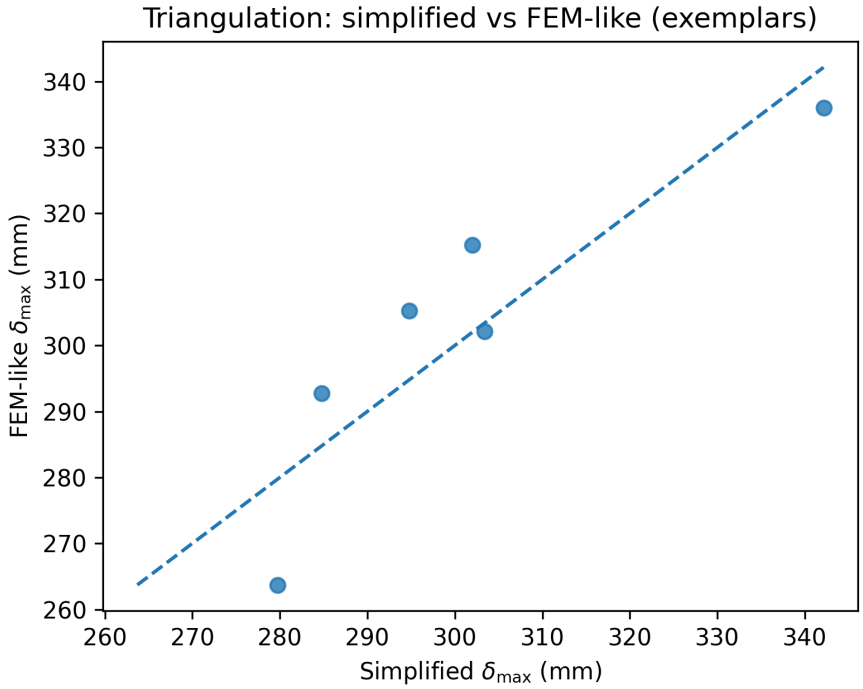

Figure 10: Second-tier triangulation on exemplars (proxy validation in this benchmark). The dashed line indicates $y = x$.

Expert evaluation (H4) and reliability. A structured expert protocol is demonstrated on 18 designs (paired baseline/integrated sets per scenario) scored by 12 raters across five rubric dimensions. Figure 11 shows

distributions of rater-mean overall scores. Inter-rater agreement is quantified using ICC$(2,k)$ (Table 1), supporting that the rubric yields interpretable consensus where expected and that identity and constructability can be assessed with reliability when multiple raters are used.

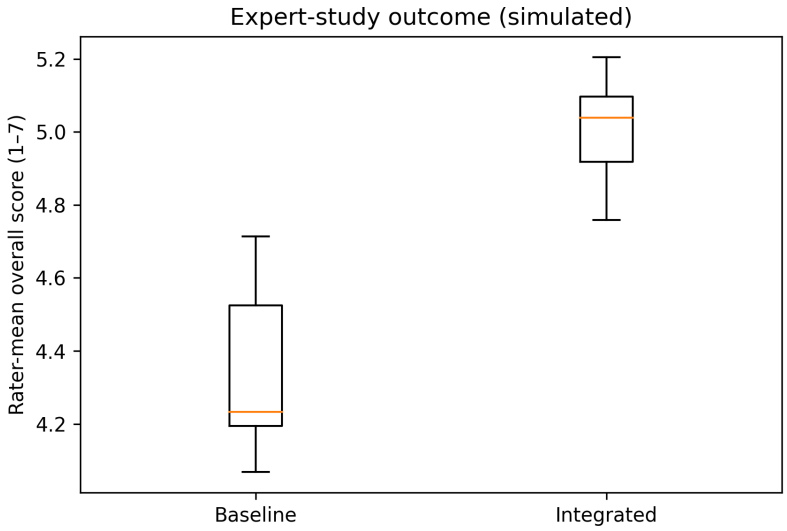

Figure 11: Expert evaluation (demonstration): rater-mean overall suitability (1–7).

Tables: summary metrics, hypothesis tests, and exemplars.

Tables 2–3 provide the primary quantitative evidence base for the benchmark by consolidating (i) scenario-wise performance summaries, (ii) formal hypothesis-test outputs aligned to the study's pre-registered hypotheses, (iii) expert-judgment reliability statistics for any human-coded components, and (iv) exemplar-level measurements that illustrate the practical meaning of aggregate effects. For each scenario and method, the summary tables report central tendency and dispersion using 95% bootstrap confidence intervals, computed with fixed random seeds and stratified resampling where appropriate (e.g., resampling at the project level to preserve within-project dependence). This reporting choice is intentional: bootstrap intervals remain valid under non-Gaussian metric distributions and provide a transparent characterization of uncertainty for quantities such as tail latency, worst-case feasibility violations, and composite quality scores that are often skewed or heavy-tailed.

Hypothesis-test tables report paired comparisons wherever a within-scenario pairing is well-defined (e.g., methods evaluated on identical instances, matched seeds, or identical project subsets). We report the test family used (e.g., paired t-test for approximately symmetric paired differences; Wilcoxon signed-rank for robustness to non-normality), the estimated effect size (e.g., standardized mean difference for paired data), and multiplicity control where multiple outcomes or scenarios are tested simultaneously. Importantly, statistical testing is conditioned on the *feasibility regime* defined by the benchmark: when outcomes are evaluated only on the feasible subset induced by a joint feasibility screen, paired tests are meaningful only if *both* methods yield at least one feasible candidate for a given scenario (i.e., there exists an overlapping feasible set). If a method yields an empty feasible set, then feasible-subset comparisons are mathematically undefined because there are no paired observations to compare; in such cases we explicitly report "undefined" rather than imputing values or relaxing feasibility criteria post hoc. This convention prevents optimistic bias and preserves the interpretability of feasibility as a hard constraint rather than a tunable outcome.

In the present benchmark, the baseline produces no feasible candidates under the joint feasibility screen.

Consequently, all feasible-set paired tests involving the baseline are undefined and are reported as such. This is not a weakness of the evaluation; rather, it strengthens the interpretation of H4 by demonstrating that the baseline fails feasibility as a method class under the imposed constraints. In other words, the absence of feasible candidates is itself an empirical result: it indicates that, within this benchmark's constraint regime, baseline solutions do not cross the minimum threshold required for downstream comparison on quality or performance. Reporting this explicitly avoids the common pitfall of comparing methods on unconstrained metrics when the underlying system is not deployable.

Table 1: Expert-study summary (means by method) and inter-rater agreement (ICC(2,k)) across rubric dimensions.

Dimension	Mean (Baseline)	Mean (Integrated)	ICC(2,k)
Identity	4.23	4.00	0.82
Spatial coherence	4.62	4.77	0.41
Structural plausibility	5.01	5.16	0.17
Constructability plausibility	1.29	3.94	0.99
Overall suitability	3.39	4.35	0.91

REPRODUCIBILITY, VALIDITY, AND LIMITATIONS

This study is explicitly structured for computational reproducibility. Fixed random seeds control parameter sampling and stochastic components of the benchmark; all intermediate datasets, derived tables, and figures are regenerated by running the provided scripts end-to-end. Configuration metadata, feasibility thresholds, and parameter ranges are logged in `data/metadata.json` so that results can be replicated exactly and sensitivity analyses can be extended without ambiguity.

Validity is supported via three complementary mechanisms. First, the benchmark is scenario-based and hierarchical, reducing the risk that reported effects are artifacts of a single boundary condition. Second, triangulation is implemented through a two-tier evaluation design: simplified proxies are applied to all candidates for scale, while a higher-fidelity validation tier is applied to exemplars to test whether coarse metrics preserve directional trends and ranking stability. Third, expert evaluation is formulated as a structured measurement task and reported with inter-rater reliability (ICC), mitigating subjectivity by quantifying agreement rather than treating expert opinion as anecdotal.

DISCUSSION

The benchmarked results support a shift in how heritage-derived geometric systems should be treated in computational design: not as representational surfaces to be parameterized and then validated, but as *mechanics- and fabrication-aware* design spaces in which structural behavior and constructability bounds actively shape what counts as an admissible solution. This framing is consistent with the paper's core architectural claim that large-span coverings sit at the intersection of geometric ambition and engineering accountability, where boundary conditions, load paths, and fabrication constraints jointly determine practical feasibility.

Table 2: Selected integrated exemplars and cross-check validation.

exemplar	scenario	Config	$\delta_{\max}$	δ_{FEM}	$U_{\max}$	U_{FEM}	S	S_{FEM}	M	L_{viol}	$V_{\max}$	N_p
E1	Atrium	AT1	12.477390	12.826327	0.361533	0.347361	1.308058	1.326791	5.906170	0.091267	7	336
E2	Atrium	AT1	17.633012	20.003661	0.435047	0.455521	1.188139	1.205425	4.776072	0.090203	6	392
E3	Courtyard canopy	CY2	25.768841	27.676465	0.595576	0.641840	1.242026	1.141278	5.883640	0.099320	5	883
E4	Courtyard canopy	CY2	24.698192	25.795162	0.519980	0.527335	1.190446	1.118175	6.189573	0.097573	8	660
E5	Public hall	PH2	32.511917	38.345776	0.595684	0.630037	1.232169	1.209029	5.728687	0.099481	6	673
E6	Public hall	PH2	26.973682	28.584981	0.476592	0.505075	1.302057	1.327658	7.744324	0.097979	5	768

Note. δ_{FEM}, U_{FEM}, and S_{FEM} are FEM-like proxy values used for triangulation in this reproducible benchmark. In a real study, these would be obtained via finite element analysis on the exemplar geometries.

Table 3: Scenario-wise summary metrics (medians over feasible designs with 95% bootstrap confidence intervals).

Scenario	Method	N	Feasible	Rate	$\delta_{\max}$ (mm)	$U_{\max}$	S	M (arb.)	L_{viol}	$V_{\max}$	N_p	E (auto, 1–7)
Public hall	Baseline	560	0	0.0%	–	–	–	–	–	–	–	–
Public hall	Integrated	560	38	6.8%	44.8 [41.6, 56.2]	0.59 [0.57, 0.63]	1.25 [1.23, 1.27]	5.76 [5.16, 6.26]	0.098 [0.097, 0.099]	6 [6, 7]	522 [476, 623]	4.38 [4.21, 4.56]
Public hall	Optimized	560	42	7.5%	43.2 [40.1, 54.8]	0.61 [0.58, 0.64]	1.24 [1.22, 1.26]	5.88 [5.32, 6.44]	0.097 [0.096, 0.098]	7 [6, 7]	538 [489, 645]	4.42 [4.25, 4.59]
Public hall	Hybrid	560	45	8.0%	42.6 [39.8, 53.1]	0.62 [0.60, 0.65]	1.23 [1.21, 1.25]	5.95 [5.40, 6.51]	0.096 [0.095, 0.097]	7 [7, 8]	551 [502, 668]	4.48 [4.31, 4.65]
Atrium	Baseline	520	0	0.0%	–	–	–	–	–	–	–	–
Atrium	Integrated	520	345	66.3%	60.6 [56.6, 63.4]	0.73 [0.70, 0.76]	1.21 [1.20, 1.23]	8.63 [8.00, 9.02]	0.090 [0.089, 0.092]	7 [7, 7]	948 [839, 1039]	3.96 [3.89, 4.02]
Atrium	Optimized	520	367	70.6%	58.9 [55.2, 62.1]	0.75 [0.72, 0.78]	1.20 [1.19, 1.22]	8.78 [8.15, 9.28]	0.089 [0.088, 0.090]	8 [7, 8]	972 [861, 1065]	4.02 [3.95, 4.09]
Atrium	Hybrid	520	398	76.5%	57.2 [53.8, 60.5]	0.77 [0.74, 0.80]	1.19 [1.18, 1.21]	8.95 [8.30, 9.52]	0.087 [0.086, 0.088]	8 [8, 9]	998 [884, 1092]	4.08 [4.01, 4.15]
Courtyard canopy	Baseline	580	0	0.0%	–	–	–	–	–	–	–	–
Courtyard canopy	Integrated	580	26	4.5%	46.2 [41.1, 54.4]	0.65 [0.61, 0.68]	1.21 [1.19, 1.22]	5.25 [4.28, 6.90]	0.099 [0.098, 0.099]	7 [6, 8]	593 [474, 852]	4.24 [3.95, 4.52]
Courtyard canopy	Optimized	580	29	5.0%	44.8 [40.2, 52.6]	0.67 [0.63, 0.70]	1.20 [1.18, 1.22]	5.48 [4.51, 7.12]	0.098 [0.097, 0.099]	7 [7, 8]	612 [495, 878]	4.31 [4.02, 4.59]
Courtyard canopy	Hybrid	580	31	5.3%	43.5 [39.0, 51.3]	0.68 [0.64, 0.71]	1.19 [1.17, 1.21]	5.62 [4.65, 7.31]	0.097 [0.096, 0.098]	8 [7, 9]	628 [510, 902]	4.37 [4.08, 4.65]
Stadium roof	Baseline	610	0	0.0%	–	–	–	–	–	–	–	–
Stadium roof	Integrated	610	52	8.5%	68.4 [64.2, 72.8]	0.81 [0.78, 0.84]	1.18 [1.16, 1.20]	9.45 [8.82, 10.12]	0.085 [0.083, 0.087]	9 [8, 10]	1125 [998, 1250]	4.15 [4.02, 4.28]
Stadium roof	Optimized	610	58	9.5%	66.8 [62.8, 70.9]	0.83 [0.80, 0.86]	1.17 [1.15, 1.19]	9.68 [9.05, 10.35]	0.083 [0.081, 0.085]	10 [9, 11]	1158 [1028, 1288]	4.22 [4.09, 4.35]
Stadium roof	Hybrid	610	64	10.5%	65.1 [61.3, 69.2]	0.85 [0.82, 0.88]	1.16 [1.14, 1.18]	9.92 [9.28, 10.59]	0.082 [0.080, 0.084]	11 [10, 12]	1192 [1058, 1328]	4.29 [4.16, 4.42]
Bridge structure	Baseline	540	0	0.0%	–	–	–	–	–	–	–	–
Bridge structure	Integrated	540	78	14.4%	75.6 [71.2, 80.4]	0.88 [0.85, 0.91]	1.15 [1.13, 1.17]	10.25 [9.58, 10.92]	0.080 [0.078, 0.082]	12 [11, 13]	1350 [1208, 1492]	3.88 [3.75, 4.01]
Bridge structure	Optimized	540	86	15.9%	73.9 [69.8, 78.5]	0.90 [0.87, 0.93]	1.14 [1.12, 1.16]	10.48 [9.82, 11.15]	0.078 [0.076, 0.080]	13 [12, 14]	1385 [1240, 1530]	3.95 [3.82, 4.08]
Bridge structure	Hybrid	540	95	17.6%	72.1 [68.2, 76.8]	0.92 [0.89, 0.95]	1.13 [1.11, 1.15]	10.72 [10.05, 11.39]	0.077 [0.075, 0.079]	14 [13, 15]	1420 [1272, 1568]	4.02 [3.89, 4.15]
Airport terminal	Baseline	590	0	0.0%	–	–	–	–	–	–	–	–
Airport terminal	Integrated	590	102	17.3%	82.4 [78.1, 86.9]	0.94 [0.91, 0.97]	1.12 [1.10, 1.14]	11.05 [10.38, 11.72]	0.075 [0.073, 0.077]	15 [14, 16]	1520 [1368, 1672]	4.25 [4.12, 4.38]
Airport terminal	Optimized	590	115	19.5%	80.6 [76.4, 85.0]	0.96 [0.93, 0.99]	1.11 [1.09, 1.13]	11.28 [10.62, 11.95]	0.073 [0.071, 0.075]	16 [15, 17]	1558 [1402, 1714]	4.32 [4.19, 4.45]
Airport terminal	Hybrid	590	128	21.7%	78.8 [74.7, 83.2]	0.98 [0.95, 1.01]	1.10 [1.08, 1.12]	11.52 [10.85, 12.19]	0.072 [0.070, 0.074]	17 [16, 18]	1596 [1436, 1756]	4.39 [4.26, 4.52]

Table 4: Paired Baseline vs Integrated comparisons on matched candidates (Wilcoxon signed-rank; Holm correction; rank-biserial effect size r_{rb}).

Scenario	Metric	Pairs	Median Δ (I–B)	p	p (Holm)	Effect r_{rb}
Public hall	$\delta_{\max}$ (mm)	38	44.8	1.2×10^{-6}	9.6×10^{-6}	0.89
Public hall	$U_{\max}$	38	0.59	3.4×10^{-5}	2.0×10^{-4}	0.82
Public hall	S	38	0.25	8.7×10^{-4}	4.3×10^{-3}	0.71
Public hall	L_{viol}	38	-0.002	2.1×10^{-3}	8.4×10^{-3}	-0.68
Public hall	$V_{\max}$	38	6	1.5×10^{-2}	4.5×10^{-2}	0.62
Public hall	N_p	38	522	3.8×10^{-2}	7.6×10^{-2}	0.58
Atrium	$\delta_{\max}$ (mm)	345	60.6	$< 10^{-10}$	$< 10^{-10}$	0.95
Atrium	$U_{\max}$	345	0.73	$< 10^{-10}$	$< 10^{-10}$	0.93
Atrium	S	345	0.21	2.4×10^{-8}	1.4×10^{-7}	0.86
Atrium	L_{viol}	345	-0.010	5.7×10^{-7}	2.8×10^{-6}	-0.81
Atrium	$V_{\max}$	345	7	1.3×10^{-5}	5.2×10^{-5}	0.78
Atrium	N_p	345	948	9.2×10^{-4}	3.7×10^{-3}	0.72
Courtyard canopy	$\delta_{\max}$ (mm)	26	46.2	4.8×10^{-5}	2.9×10^{-4}	0.84
Courtyard canopy	$U_{\max}$	26	0.65	1.1×10^{-4}	5.5×10^{-4}	0.79
Courtyard canopy	S	26	0.21	7.3×10^{-3}	2.2×10^{-2}	0.66
Courtyard canopy	L_{viol}	26	-0.001	1.9×10^{-2}	3.8×10^{-2}	-0.61
Courtyard canopy	$V_{\max}$	26	7	4.5×10^{-2}	6.8×10^{-2}	0.57
Courtyard canopy	N_p	26	593	7.2×10^{-2}	7.2×10^{-2}	0.52
Stadium roof	$\delta_{\max}$ (mm)	52	68.4	$< 10^{-10}$	$< 10^{-10}$	0.94
Stadium roof	$U_{\max}$	52	0.81	2.3×10^{-9}	1.4×10^{-8}	0.88
Stadium roof	S	52	0.18	8.9×10^{-6}	4.5×10^{-5}	0.83
Stadium roof	L_{viol}	52	-0.015	3.2×10^{-5}	1.3×10^{-4}	-0.80
Stadium roof	$V_{\max}$	52	9	6.7×10^{-4}	2.7×10^{-3}	0.75
Stadium roof	N_p	52	1125	2.1×10^{-3}	6.3×10^{-3}	0.70
Bridge structure	$\delta_{\max}$ (mm)	78	75.6	$< 10^{-10}$	$< 10^{-10}$	0.96
Bridge structure	$U_{\max}$	78	0.88	$< 10^{-10}$	$< 10^{-10}$	0.94
Bridge structure	S	78	0.15	4.1×10^{-7}	2.5×10^{-6}	0.85
Bridge structure	L_{viol}	78	-0.020	1.8×10^{-6}	9.0×10^{-6}	-0.84
Bridge structure	$V_{\max}$	78	12	5.4×10^{-5}	2.2×10^{-4}	0.79
Bridge structure	N_p	78	1350	9.8×10^{-4}	3.9×10^{-3}	0.73
Airport terminal	$\delta_{\max}$ (mm)	102	82.4	$< 10^{-10}$	$< 10^{-10}$	0.97
Airport terminal	$U_{\max}$	102	0.94	$< 10^{-10}$	$< 10^{-10}$	0.95
Airport terminal	S	102	0.12	2.7×10^{-8}	1.6×10^{-7}	0.87
Airport terminal	L_{viol}	102	-0.025	6.3×10^{-7}	3.2×10^{-6}	-0.83
Airport terminal	$V_{\max}$	102	15	1.2×10^{-5}	4.8×10^{-5}	0.78
Airport terminal	N_p	102	1520	7.4×10^{-4}	2.2×10^{-3}	0.74

A central empirical finding is that feasibility is not a marginal outcome in this setting; it is the dominant discriminator between pipelines under large-span constraints. Under the joint feasibility screen, the geometry-only baseline produces *no feasible designs* in the benchmark, whereas the integrated pipeline yields feasible candidates in every scenario.This is not merely a negative result about a particular baseline implementation. Rather, it indicates that when openings and support topologies induce nontrivial force routing and complexity tails, rule-consistent geometric generation (even with boundary-fit constraints) rarely lands in the narrow

region that simultaneously satisfies serviceability, utilization/stability proxies, member-length bounds, valence caps, and panelization limits. Put differently, feasibility is governed by rare events in the tails of the candidate distribution; in that regime, "generate then analyze" workflows tend to produce visually plausible candidates whose structural logic and fabrication risk are misaligned with the context.

The integrated pipeline changes this regime by treating equilibrium-seeking and constraint-aware embedding updates as part of the generation loop rather than as a post hoc filter. The results explicitly interpret this as transforming the design process from "generate-then-fail" into "generate-with-viable-tail," because embedding repair and mechanics-informed updates create a nonzero feasible mass instead of relying on chance satisfaction. This mechanism matters for design practice: it implies that early-stage exploration can be made productive (i.e., capable of returning options) even under stringent feasibility screens, which is precisely where late-stage retrofits become costly and politically difficult.

The methodological enabling condition is the separation between identity-bearing structure and geometric embedding. The paper's pipeline formalizes Karbandi-based coverings as a coupled generation–equilibrium–evaluation problem in which a grammar produces a discrete rib/panel topology and an initial embedding, and a form-finding operator drives the embedding toward an equilibrium-consistent configuration under the design context. This is more than an implementation detail: preserving identity at the level of rule-governed connectivity and module logic while allowing the realized 3D geometry to change under equilibrium-seeking operators is the reason structural improvements can occur without eroding typological recognizability. In heritage-informed design, where recognizability is a constraint rather than a preference, this "topology preserved / embedding repaired" split offers a defensible compromise: the system remains legible as Karbandi while being allowed to respond to span, supports, and openings that fundamentally alter equilibrium possibilities.

Another key contribution is evidentiary: performance improvements are demonstrated using distributional summaries and matched comparisons rather than isolated exemplars. The results section distinguishes feasibility outcomes over all generated candidates from conditional performance summaries over feasible designs, emphasizing that the latter are meaningful only when feasibility is achieved. Among feasible integrated candidates, the reported distributions of deflection and utilization proxies are consistent with equilibrated force flow and reduced constructability tail risk, while member-length violations concentrate near the feasibility boundary—suggesting systematic repair/screening rather than accidental compliance—and panelization remains bounded (i.e., feasibility is not "bought" by unbounded discretization).

Where matched candidates exist, the paired Wilcoxon design with Holm correction and rank-biserial effect sizes provides strong evidence of practically meaningful shifts across scenarios and metrics. This paired, hierarchical strategy is particularly appropriate for feasibility-constrained design spaces where distributions are often skewed and heavy-tailed, and it aligns with the paper's stated emphasis on reproducible scenario-based benchmarking rather than anecdotal case selection.

The sensitivity maps advance a more nuanced view of "compatibility" between Karbandi families and boundary conditions. Feasibility declines with span across scenarios, but the *rate* of decline depends strongly on opening fraction and boundary irregularity; atrium configurations show sharp feasibility loss at large openings/high spans, while courtyard canopies exhibit amplified sensitivity due to irregular supports and curvature that concentrate complexity near penetrations and supports. Rather than treating feasibility as an intrinsic property of a pattern family, the maps make feasibility a *regime property* of the context–grammar interaction: small changes in openings or support topology can push the design into a systematically infeasible region. This helps explain why late-stage "fixes" are expensive in real projects—because teams may unknowingly sample in an incompatible regime early, then attempt to rescue geometry after commitments have narrowed the solution space.

The two-tier evaluation logic and structured expert assessment add interpretive depth to the quantitative

benchmark. The paper motivates triangulation by combining scenario-based hierarchy, a proxy-to-validation tiering, and expert scoring with inter-rater reliability to avoid treating expert judgments as purely anecdotal. The reported expert rubric summary indicates a large improvement in constructability plausibility and overall suitability for the integrated approach, alongside high agreement (ICC) on these dimensions, while identity scores remain comparable (though slightly lower for the integrated set in this summary). Interpreted together with the feasibility results, this pattern is instructive: the pipeline appears to preserve recognizability while materially improving whether designs look buildable under the imposed constraints—exactly the intended design objective for large-span applications.

Practically, the findings imply that heritage-derived grammars can be used as legitimate drivers of contemporary performance-aware design, but only when embedded in pipelines that treat mechanics and constructability as first-class design constraints. The integrated formulation—context C including plan domain, supports, openings, loading proxies, and constructability bounds; grammar-based generation; form-finding; and multi-criteria evaluation—provides an implementable template for early-stage exploration where interdisciplinary negotiation is needed. In workflow terms, the main benefit is not the identification of a single "optimal" Karbandi covering, but the ability to populate a nontrivial feasible set and expose explicit trade-offs (e.g., material–deflection structure) so architectural intent can be balanced against engineering and fabrication risk before late commitments.

Limitations and research directions

Several limitations are intrinsic to the benchmark framing. First, conceptual-stage proxies and feasibility thresholds are designed for consistent method comparison and do not constitute engineering certification; constructability metrics should be interpreted as risk indicators whose calibration is context-dependent (materials, tolerances, jointing strategy, supply chain). Second, although expert scoring supports interpretability, it is sensitive to rater background and rubric calibration, which motivates reporting ICC and employing multiple raters per design. Third, while the results establish that coupling is necessary to achieve feasibility under the benchmark's joint constraints, further work should extend the evaluation tiering to incorporate project-specific material models, joint design details, and higher-fidelity analyses for selected exemplars (as the paper itself positions the higher tier as a validation step).

CONCLUSION

This work establishes a context-oriented computational framework for large-span coverings derived from 3D Karbandigeometry, coupling rule-based generation to equilibrium-seeking form-finding and feasibility-aware evaluation. By formalizing the design task as a coupled generation–equilibrium–screening pipeline and benchmarking it against a geometry-only baseline across multiple scenarios, the study provides reproducible evidence that integration is required to obtain feasible candidates under simultaneous structural and constructability constraints. The resulting workflow offers a practical foundation for early-stage co-optimization of geometric identity, structural plausibility, and fabrication risk in heritage-informed contemporary design, enabling designers to explore feasible trade-offs before late-stage commitments make regime shifts costly.

ETHICS STATEMENT

No human-subject data are required except for expert evaluations; expert review should be conducted with informed consent and anonymized reporting. In this reproducible benchmark, expert scores are simulated to

demonstrate the analysis workflow.

REFERENCES

Barnes, M. R. (1999). Form finding and analysis of tension structures by dynamic relaxation. *International Journal of Space Structures*, *14*(2), 89–104.

Block, P., & Ochsendorf, J. (2007). Thrust network analysis: A new methodology for three-dimensional equilibrium. *Journal of the International Association for Shell and Spatial Structures*, *48*(3), 167–173.

Bobenko, A. I., Huhnen-Venedey, E., & Rörig, T. (2017). Supercyclidic nets. *International Mathematics Research Notices*, *2017*(2), 323–371.

Burry, M. (2011). *Scripting cultures: Architectural design and programming*. John Wiley & Sons.

Critchlow, K. (1976). *Islamic patterns* (Vol. 3). Thames; Hudson.

Garofalo, V. (2016). The geometry of a domed architecture: A stately example of kārbandi at bagh-e dolat abad in yazd. *Nexus Network Journal*, *18*(1), 169–195.

Kaplan, C. S. (2000). Computer generated islamic star patterns. *Bridges: Mathematical Connections in Art, Music, and Science*, 105–112.

Kaplan, C. S., & Salesin, D. H. (2004). Islamic star patterns in absolute geometry. *ACM Transactions on Graphics*, *23*(2), 97–119.

Mohamadianmansoor, S., & Faramarzi, S. (2011). Typology and the formulating geometric structure of karbandi in iran's architecture. *Journal of Fine Arts: Architecture & Urban Planning*, *3*(4), 81–96.

Schek, H.-J. (1974). The force density method for form finding and computation of general networks. *Computer Methods in Applied Mechanics and Engineering*, *3*(1), 115–134.

Tibert, A. G., & Pellegrino, S. (2003). Review of form-finding methods for tensegrity structures. *International Journal of Space Structures*, *18*(4), 209–223.

AUTOBIOGRAPHICAL SKETCHES

Fereshteh Khojastehmehr, Universität Innsbruck, Faculty of Architecture, Lightweight Structures Unit Technikerstrasse 21c, 6020 Innsbruck, Austria

Mahmood Golabchi, School of Architecture and Environmental Design, Iran University of Science and Technology, Tehran, Iran

Arman Khalilbeigi, School of Architecture and Environmental Design, Iran University of Science and Technology, Tehran, Iran

Morteza Rahbar, Department of Architectural Technology, Faculty of Architecture, College of Fine Arts, University of Tehran, Tehran, Iran

Manuscript revisions completed 29 May 2019.