Local QEF Displacement Bounds in Implicit CSG Supplementary Material

Jeremy Pei

This supplement retains descriptive scans, auxiliary tracks and complete saved timing pairs. Sections S1–S4 retain the historical evidence; Section S5 retains the untimed unit-structure control; Section S6 retains the fixed-mesh surface-audit runs. Their original timing windows remain separate. Section S7 retains stored-data correctness obligations and adds a separately recorded endpoint/exception hardening check. Section S7.6 distinguishes saved-record interfaces and records a separate import repair and its remaining trust boundaries. Section S8 reports the matched-input numerical control, and S9 documents the still-unmet engineering interface and locked recomputation protocol. None of these changes recertifies historical field budgets or adds an end-to-end CAD speed claim.

S1 Full policy grid and confirmation interpretation

The local confirmation protocol was hashed before its nine tasks were extracted, but it was not publicly preregistered. The descriptive grid is

$$\kappa \in \{0.03, 0.06, 0.12, 0.24, 0.48, 0.75, 0.95, \text{None}\}, \qquad \tau_C/h_C \in \{0.10, 0.25, 0.50\}.$$

`None` removes only the optional normal policy, not $g < 1$ or the geometric/mathematical guards. Four methods (normal only, minimum/λ, spectral, interval) give 96 configurations. Main contrasts were $\kappa = 0.12, 0.24$ at $\tau_C = 0.5h_C$. Full taskwise records retain empty admissions and Unknown outcomes; an empty set has undefined conditional risk, not a zero-risk operating point.

Every selected leaf's source-cell phase width $h_C\|k_j\|_1$ in this confirmation lies between 3.2573 and 3.5258, exceeding the period π of $|\cos|$. Thus the envelope reaches full amplitude and the per-cell maximum gradient budget is at least 0.132. Rejection at $\kappa = 0.12$ is analytically predictable. Thresholds at least 0.24 give the same eligible population. The relaxed contrast was fixed before extraction, but the design cannot resolve fine effects of those larger optional thresholds. The negative outcome is specific to this design, not proof that QEF gates are universally useless. The stopping rule was applied rather than adding similar tasks.

The separate 50,000 algebraic systems are historical fixed-seed row problems, not additional CAD shapes. Their 23,577 denominator gains and 13,934 extra admissions at tolerance 0.5 test the mechanism under different row, regularization and uncertainty distributions. They do not isolate geometry as the sole cause of the contrast with CSG cells.

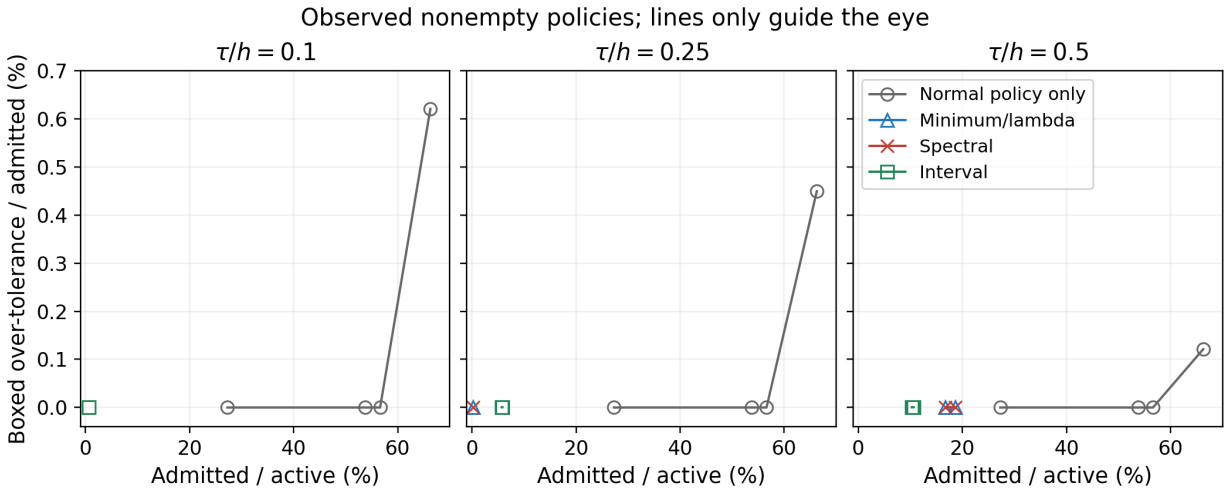

Figure S1: All nonempty historical policy configurations. Risk is the finite fraction of admitted boxed candidate–reference displacements above tolerance. Markers are tested configurations; connecting lines do not interpolate valid policies. These historical sweeps are exploratory.

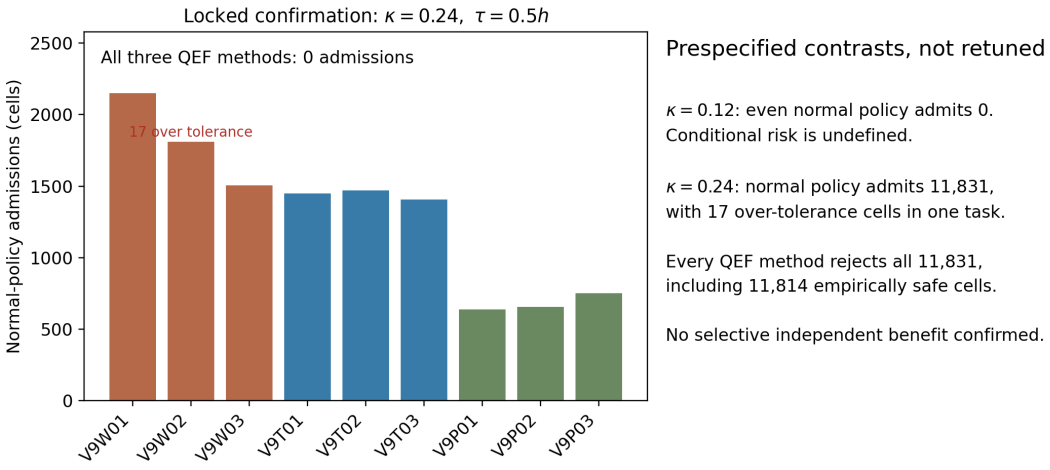

Figure S2: All nine fixed confirmation outcomes. The 17 observed QEF failures occur in one thin-wall variant, V9W02. All QEF tests reject every eligible cell, including 11,814 audit-safe candidates. These tasks are already known stress cases in the later verification-cost study, not new independent confirmation.

S2 Auxiliary STEP, reconstruction and producer tracks

S2.1 Historical STEP diagnostic

The historical real-CAD track contains 528 STEP inputs, 511 no-healing mesh admissions, 122 selected DC runs, and 118 main-protocol completions. These counts describe different stages and are not interchangeable. We retain its frozen records as retrospective diagnostic evidence. In v8, all eleven NIST inputs available in the recorded source ledger are bundled with their original hashes, attribution, project-specific source, and a portable replay command. The product-specific NIST permission statement permits their reuse. The other 517 historical paths were unavailable in the audited mirror; this is a local-availability finding, not a redistribution-rights determination.

The standalone NIST replay completes all eleven models and produces 20,387 active cells, of which 3,010 pass the oracle-assisted diagnostic test (14.76%), with zero observed bound violations. Its fixed settings are grid resolution 32, regularization 0.1, and tolerance 0.1 cell widths. The experiment starts from the final STEP solid and estimates errors using a mesh reference. It does not recover a CSG tree or supply certified upstream field bounds. Its pass rate is therefore reported separately from the analytic-CSG admission study. The eleven-source replay requires OpenCASCADE, a compiler and CMake. Its input hashes, attribution, source and replay commands are retained in the historical geometry archive; the integration did not rebuild it. The absent 517 historical inputs limit what a fresh recipient can reproduce and are not silently counted as bundled source data. Neither reference-derived budgets nor successful source reconstruction establish independent upstream guarantees.

S2.2 External reconstruction adapters

Official DCSDD and GCP with screened Poisson runs on two compatible exact-SDF grids are retained as historical build/adapter checks. They use different input access and output objectives from uncertain-row QEF verification. The runs are not a same-information CSG comparison or a state-of-the-art ranking. They are not pooled with the 22-task verification-cost workload.

S2.3 Separate sampled-grid pilot

The bounded sampled-grid pilot uses exact rational arithmetic for quantized samples, trilinear interpolation errors, component gradients and interpolant Lipschitz constants under supplied valid C^2 Hessian bounds. Its five fixed plane/sphere cases contain 4,408 queries: 3,888 valid interiors and 520 explicit Unknown outcomes at interfaces, outside the domain or without the regularity premises. Its saved deterministic replay reports no observed enclosure violation; the manuscript integration did not rerun it.

This is producer-level feasibility, not an integrated Hermite/QEF pipeline. A C^2 field need not be an SDF. A gradient-error budget supplies neither unit reference normals nor positive reach; interfaces are rejected and witness-ball domain extension remains unresolved. These records cannot support claims about arbitrary sampled or neural fields, CFD accuracy, peak-memory improvement or complete-pipeline acceleration.

S3 Saved paired timings and full cost disclosure

The following historical tables were generated from the locked verification-cost JSON, excluding warmups, superseded scale records and aborted runs from the main aggregates. For each repeat, task walls are summed within the specified workload before computing a group median. Costs use the fixed host and declared measurement window, not per-cell universal prices.

Cache-fill amortization in the table is descriptive for the five recorded hot service calls, not a predicted application workload. Excluded costs include post-return cache/container destruction and garbage collection, audit comparison, report construction and writing. Upstream field production, Hermite extraction and surface completion are also excluded. Consequently a hot hit must not be quoted as the total cost of establishing the guarantee from new geometry.

The difference between the merged B and C cold medians is about 42.63 seconds. This is a difference of two group medians, not the median of paired differences, a marginal price per cell, or a downstream break-even threshold. A/B share an acceptance requirement; C does not provide the one additional spectral admission; D has a different, overlapping acceptance set. These are distinct comparisons.

Table S1: All five observed group-total pairs underlying the main-text RQ5 cost table. Positive $A - B$ means less time for B. No repeat is an independent geometric model or statistical participant.

Workload	Repeat	A (s)	B (s)	$A - B$ (s)
Cold: historical (13)	1	31.1879	23.6689	+7.5189
	2	31.4017	23.6530	+7.7487
	3	31.2753	23.5670	+7.7083
	4	31.3503	23.6768	+7.6735
	5	31.3336	23.5574	+7.7762
Cold: known stress (9)	1	26.4193	26.4686	-0.0493
	2	26.5425	26.5786	-0.0361
	3	26.5110	26.4666	+0.0444
	4	26.4842	26.4125	+0.0717
	5	26.3899	26.4636	-0.0738
Cold: combined (22)	1	57.6071	50.1375	+7.4697
	2	57.9442	50.2316	+7.7126
	3	57.7863	50.0336	+7.7527
	4	57.8346	50.0893	+7.7453
	5	57.7235	50.0210	+7.7025
Hot: combined (22)	1	0.6782	0.5070	+0.1712
	2	0.6524	0.5184	+0.1339
	3	0.3904	0.5386	-0.1482
	4	0.3981	0.7722	-0.3741
	5	0.5205	0.5163	+0.0042

Table S2: Preparation-inclusive bookkeeping for the combined 22-task workload (seconds), reconstructed from the same saved run. C is cheap-only; D is interval-only. Different admission sets prevent treating C/D as the same service as A/B.

Method	Input preparation	Cold validation median	Input+cold one service	Hot fill incl. creation	Hot service five requests	Amortized per request
A	11.5725	57.7863	69.3588	57.7334	71.9453	14.3891
B	11.5725	50.0893	61.6618	50.1612	64.5863	12.9173
C	11.5725	7.4639	19.0364	7.5737	21.7262	4.3452
D	11.5725	7.9899	19.5624	8.3649	22.3555	4.4711

Input preparation is the one shared measured load/hash/prerequisite/snapshot phase. 'Input+cold' adds it to the median of each formal cold total plus that repeat's measured cache-container creation. Hot fill is each strategy's own hot warmup validation plus container creation; B does not prefill spectral results for early-accepted cells. 'Hot service five requests' includes input preparation, that fill and all five subsequent hot totals; the last column divides this service total by five, excluding warmup as a serviced request. Result/container teardown outside the timed batch was not separately measured, so these are not full-lifecycle or end-to-end CAD timings.

S4 Study chronology, archive layers and checks

Internal version names identify frozen archives, not additional independent experiments. The main text instead names populations by their scientific role. The map below preserves the chronology used to interpret the evidence.

Archive alias	Scientific role and interpretation
v9	Historical CSG and algebraic studies; the nine locally predeclared confirmation tasks were generated after the initial exploratory studies.
v10–v11	Later verification-cost protocol on the same 22 known tasks; D01 scale correction and superseded/interrupted runs remain separate.
v12	Untimed unit-structure control on the same declarations; no new timing claim.
v13 / Stage09–11	Fixed-mesh distance audit, including its failed predecessor; this does not change Hermite rows or demonstrate QEF-guided reconstruction.
v15–v16	Restricted stored-data realization and integration; 200 small-grid records are software fixtures, not new CAD-model replications.
v13.2–v13.3	Known-data common-box comparison, engineering preflight and editorial integration; not a rerun of historical geometry or timing.
v13.3.1–v13.3.2	A premise clarification, followed by endpoint and exception-state hardening with bounded software regressions; no new scientific population or revision of historical results.

The companion evidence bundle retains the two outer archives v13 and v16. The former nests v12, v11 and v9 and includes the portable v10 cost layer; the latter includes the restricted v15 material and the runtime needed for its stored fixtures. The companion manifest records hashes and extraction paths. The historical geometry layer is materialized into a new relative-path directory when needed, avoiding duplicated large source trees.

Source-identity checking, summary regeneration, software regression, scientific re-execution and independent replication are different actions. Each release receipt separates retained historical evidence from new actions. Timing reruns are optional and must write new records, not replace the fixed reported times. D01's correction and superseded/interrupted runs remain navigable. A local evidence index or read-only assistant may aid discovery, but neither is required to inspect or reproduce the core scientific evidence.

Assistant software tests and public development examples remain separate. No human workflow, model-based question-answering accuracy or productivity evaluation is part of this paper.

The auxiliary surface audit retains 27 predecessor records and 36 later records on three fixed mesh pairs. The saved audit reports matching repeat proofs for all 36 records and 1,122,136 exact component comparisons on 12 distinct proofs. These are software reverifications, not independent human replication. Reported areas still use floating-point triangle areas. The oldest protected-workspace check also depends on 102 unbundled build, cache and receipt files; its original complete-workspace invocation is not claimed portable. This restriction is listed separately from the available saved-result and later partition-proof checks.

S5 Untimed unit-structure control

The v12 control uses the original 22-task declaration population and changes only the Hessian error budget; it is a post-review check on known data. For the same certified spectral shift $\underline{\ell}$, it computes

$$\eta_H^{\mathrm{unit}} = (\sqrt{m} + \widehat{\beta})\eta_N, \qquad d_{\mathrm{unit}} = \max(\lambda, \underline{\ell} - \eta_H^{\mathrm{unit}}),$$

where $\widehat{\beta}$ is an outward Frobenius upper bound for the actual binary64 observed rows. The same legacy outward numerator is divided by the new denominator; taking the minimum with the legacy bound preserves the fallback. Exact reference unit normals remain a semantic premise.

Table S3: Untimed unit-row Hessian control on the fixed v10 declarations. Smaller bound compares against the legacy spectral bound; both use the same numerator and verified shift.

Population	Eligible	Smaller bound	Legacy / unit accepts	Added
Historical	14,021	28	3,945 / 3,945	0
Known stress	11,831	2	0 / 0	0
Combined	25,852	30	3,945 / 3,945	0

Table S4: Taskwise fixed-declaration control. All Hessian budgets decrease. D1/D2 count legacy/unit denominators above λ; B counts smaller final bounds.

Task	Eligible	D1/D2	B	Legacy/unit accepts	Added
D01	808	13/14	14	808/808	0
D02	1322	0/0	0	385/385	0
I01	406	0/0	0	290/290	0
I02	457	0/0	0	169/169	0
M01	1876	0/0	0	0/0	0
M02	1339	2/2	2	308/308	0
M03	1136	12/12	12	791/791	0
M04	303	0/0	0	0/0	0
M05	2044	0/0	0	0/0	0
M06	600	0/0	0	0/0	0
U01	1009	0/0	0	969/969	0
U02	1446	0/0	0	2/2	0
U03	1275	0/0	0	223/223	0
V9W01	2151	0/0	0	0/0	0
V9W02	1810	0/0	0	0/0	0
V9W03	1507	0/0	0	0/0	0
V9T01	1448	0/0	0	0/0	0
V9T02	1469	0/0	0	0/0	0
V9T03	1406	0/0	0	0/0	0
V9P01	637	2/2	2	0/0	0
V9P02	654	0/0	0	0/0	0
V9P03	749	0/0	0	0/0	0

All 25,852 eligible records are retained with input hashes, budgets, denominators, bounds, and decisions. Original per-cell bounds are recomputed and checked for exact equality before interpreting the new control. The Hessian budget is smaller on every record; only 30 final bounds become smaller and none crosses a new admission threshold. The 3,945 accepted identifiers are unchanged. The one new denominator above λ is in D01, whose candidates were already accepted.

An independent aggregation script checks the per-cell records, reconstructs the table totals, and compares new bounds with the stored boxed candidate–reference displacements using the previously disclosed audit slack. This is not a fresh exact-geometry reconstruction or an upstream certificate. Source/unit tests include rational matrix identities and path handling; their cases are not added to the research population. No new official timing was performed, so the old cost tables do not measure this control.

The public v12 H1 entry point resolves relative output paths under the materialized v10 runtime root before invoking the unchanged old verifier. It rejects outside-root paths, existing outputs and partial/failed runs. Both the original faulty relative-path command and its numerical source remain preserved in the frozen v11 archive; the v12 guide supersedes that entry-point instruction without rewriting historical evidence.

S6 Fixed-mesh surface-audit development

This layer is separate from the analytic CSG/QEF tasks and the historical STEP diagnostic. It reuses a previously produced reference/output mesh pair without running a CAD kernel or extractor. The target is the frozen Stage08 repaired P02 mesh, not the earlier pre-repair output. No CGAL is used.

S6.1 Summary retained from the former main-text auxiliary result

This is a coverage–resource comparison, not a QEF recomputation experiment. W verifies 98.5866% of P02's numerical reference area and MW verifies 98.6338%, compared with E's 87.6666%, at the same 200,000 global-query cap. MW gains only 0.0472 percentage points over W while roughly doubling explicit local point–triangle tests. W/MW take 29.64/30.34 s versus E's 23.57 s, but cover different amounts of surface. Neither passes the primary whole-surface gate. The remaining MW Unknown area is 269.19 mm^2 and intersects 950 of the 3,778 original reference faces. These may include important features; area alone does not establish their insignificance.

Table S5: P02 fixed-mesh audit at 200,000 global queries for each method. Area fractions are numerical weights of strictly accepted patches. Full CLI time is the median of three repeats; it excludes the separate independent proof-log audit and upstream extraction.

Method	Covered (%)	Unknown (%)	Local tests (10^6)	CLI (s)
E: centre	87.6666	12.3334	0.200	23.57
F: face + reset	91.3221	8.6779	4.942	26.34
W: + walk	98.5866	1.4134	16.541	29.64
MW: + four pieces	98.6338	1.3662	34.553	30.34

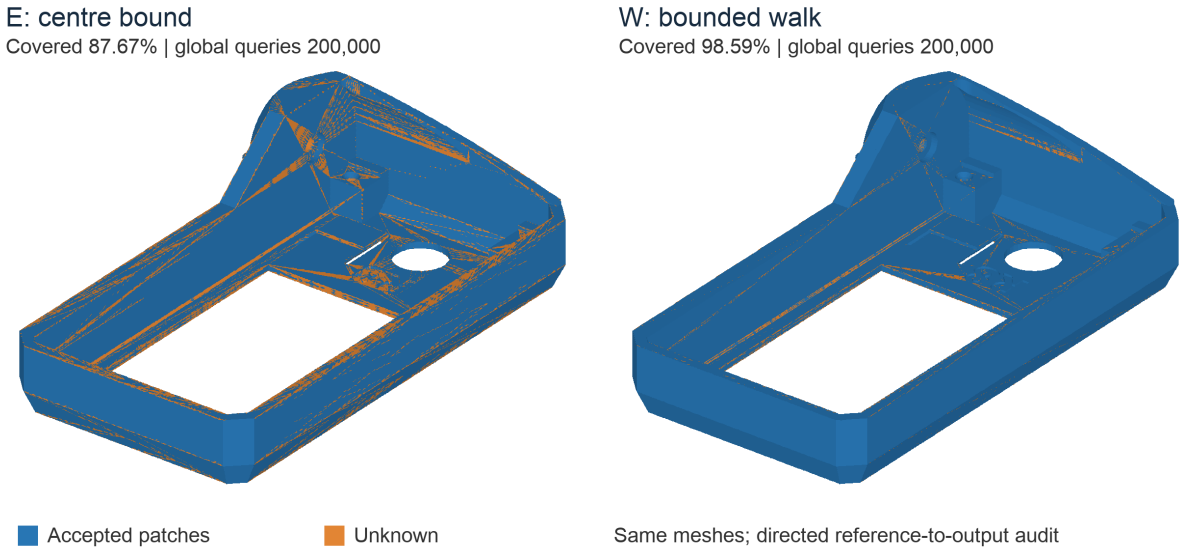

Figure S3: Historical P02 fixed-mesh audit. Blue and orange show accepted and Unknown patches in the saved E and W partition proofs; hidden surfaces are occluded in this view. Both use the same output mesh. This result is not evidence of a QEF-guided change in reconstruction. Source: Prusa Research, Original Prusa MINI, MINI-display-box; revision and license are retained.

S6.2 Object, arithmetic and patch proofs

Let $P = \mathrm{conv}\{p_0, p_1, p_2\}$ be a source patch and let M be the union of the actual target triangles. Put $d(x, M) = \inf_{y \in M} \|x - y\|$. For $c = (p_0 + p_1 + p_2)/3$ and $r = \max_i \|p_i - c\|$, any $q \in M$ gives

$$\sup_{x \in P} d(x, M) \le \|c - q\| + r.$$

Indeed, convexity gives $\|x - c\| \le r$, and the triangle inequality then applies. No nearest-point property is required of q.

For a single actual target triangle $S \subset M$, choose $q_0, q_1, q_2 \in S$. With $x = \sum_i \lambda_i p_i$, where $\lambda_i \ge 0$ and $\sum_i \lambda_i = 1$, the point $y = \sum_i \lambda_i q_i$ remains in S, and

$$d(x, M) \le \|x - y\| \le \sum_i \lambda_i \|p_i - q_i\| \le \max_i \|p_i - q_i\|.$$

This does not permit combining points on unrelated target faces into an unverified convex hull. Splitting P at its three edge midpoints gives four source subtriangles P_j whose union is P. Each may have a different actual target face, but all four upper bounds must pass; taking the maximum then proves the claim on P. The implementation either accepts the whole four-piece certificate or continues binary subdivision of the parent. It does not count a partly accepted four-piece candidate as an accepted parent.

Source barycentric coordinates are exact dyadic midpoint coordinates, with depth at most 24. Target witnesses use nonnegative integer barycentric weights summing to 2^{30} on the recorded actual target triangle. Binary64 input coordinates are interpreted as exact rational numbers, rather than as enclosures of an unknown ideal CAD surface. Squared paired distances are compared by integer arithmetic. For the centre bound, writing $a = \|c - q\|^2 \ge 0$ and $b = r^2 \ge 0$, with $\tau \ge 0$, gives an exact test of $\sqrt{a} + \sqrt{b} \le \tau$: require $R = \tau^2 - a - b \ge 0$ and $R^2 \ge 4ab$. The rational factor $1/3$ in c is preserved, not silently rounded before this test. No epsilon is added to the acceptance threshold.

The independent proof-log checker rebuilds each source root, verifies parent/child geometry and unique ownership, and checks the exact dyadic partition mass for every root. It then rechecks every accepted leaf's witness membership and inequality. Area percentages use floating-point root areas and exact dyadic subdivision fractions; they are not outward enclosures of total area. Whole-surface success requires every leaf to pass, not a rounded display of 100%. An unaccepted leaf is Unknown unless a separately recorded global numerical query supplies a numerical exceedance; the latter is not a certified distance lower bound. No such exceedance is recorded in the current P02 runs.

S6.3 Frozen cases, schedule and search

The controls are a thin slab and separated slabs, both at $h = 0.3125$ mm and $\tau = 0.15625$ mm. The P02 grid scale is $h = 0.8817187500000547$ mm and its parsed threshold is 0.44085937500002736 mm. Its numerical reference area is 19702.980065384 mm^2; its reference/target have 3,778/431,350 triangles. The source is Prusa Research's Original-Prusa-MINI repository, `STEP/PRINTED PARTS/MINI-display-box.stp`, revision `853bc30c4b10190f1d669ed6d0a567e333c28f21`. Model attribution, the repository's GPLv3 license and exact input hashes are retained. This engineering example is not a new sample from the historical QEF task populations.

The deterministic scheduler processes largest-area patches first, breaks ties by creation ID, and uses longest-edge midpoint bisection with fixed edge-index ties. All variants share 200,000 global queries, 1.5 million nodes, depth 24, batches of 512, and soft search limits of 240 s and 1,800 MiB for the sum of self/native separate peak reports. A parent with cached witnesses and radius larger than 2τ is subdivided before local testing; E instead uses its centre-bound radius screen. In F/W/MW, root patches obtain an initial global witness; E may first subdivide a root whose radius exceeds τ. Returned batches are finished at the global cap; the remaining queue is explicitly Unknown.

F evaluates same-face bounds using a small candidate group. After any failed strict check, both children's caches are invalidated and they must obtain fresh global candidates. W further descends quantized centre-witness distance along selected edge-neighbor links for at most 64 strict decreases of the floating-point proposal score. For each edge, the smallest-ID other incident face is stored: this is a bounded selected-adjacency graph, not the complete neighborhood of a nonmanifold edge. The final neighborhood and an optional global-fallback neighborhood contain at most eight candidate faces. The lowest proposed whole-face bound, with its actual face and witness weights, is retained across the pre-walk, post-walk and fallback evaluations; final acceptance is checked exactly. MW applies its four subtriangle tests to the same

final candidate group; it does not run four independent walks. Local walks need not find a nearest face, and local witness distances never certify a failure lower bound. Every explicit point–triangle test is counted.

The three fixed variant orders are E/F/W/MW, F/W/MW/E and W/MW/E/F for each case. The first repeat retains its full node proof; subsequent identical proofs retain their digest and full counters. A mismatch would retain the new proof and fail integration. None occurs. This is local pre-outcome locking, not public preregistration or blind confirmation. The secondary 95% gate is a development criterion, not a confidence level or an engineering safety standard.

S6.4 Failed predecessor and measured ablations

Stage10 ran E, cached-face F and four-piece M three times on the same three cases. Its P02 coverages were 87.6666%, 91.3221% and 91.4055%. F/M each left 738 Unknown leaves on the thin-slab control: coverage was 99.9999681301%, not complete acceptance. All 2,956 strict rejections per such run are retained. Rejected candidates were inherited after splitting, so the same unsuitable cached face could survive to the depth cap. This was an incomplete candidate-refresh policy, not permission to relax the inequality. The failed gate motivated the separately locked Stage11 cache reset and bounded walk; no Stage10 result is overwritten.

In Stage11 every control leaf is accepted by every variant. On P02, all variants stop at the global-query cap rather than the depth or node cap. The complete partition is retained even when some leaves are Unknown. W/MW respectively make 1,466,472/1,464,905 local improving walk moves. The small incremental gain from MW does not justify claiming that four-piece certificates are always the best engineering choice. The historical Stage09 four-way FIFO, longest-edge FIFO, longest-edge area, same-face area and prescreen-area variants remain in its unchanged archive; its 10.14% headline is not used as the sole contemporary comparator.

S6.5 Costs, residual area and verification limits

Full CLI wall time includes parsing, native setup, inline exact checks, partition checks, output serialization and process lifetime. Repeat zero compresses its proof whereas identical later repeats write a digest, so serialization cost is not identical across repeats. The three times are descriptive host measurements, not controlled cold-cache timings or independent-model statistics. The separate 78.08 s Stage11 proof-log audit, fresh compilation/replays, reports and all upstream CAD/extraction work are excluded. Global AABB traversal internals are not instrumented as explicit local tests. Peak-sum memory combines separate process peaks; it is not an observed simultaneous system-memory maximum. There is no memory-reduction or full-pipeline speedup claim.

W leaves 34,975 Unknown patches on 958 source faces; MW leaves 33,892 on 950 source faces. The three largest MW per-root Unknown contributions are 11.92, 11.23 and 8.90 mm^2 (zero-based reference faces 3523, 3522 and 2660). Unknown is not confined to a few faces. The artifact reports depth and spatial summaries, but its root-edge connectivity is only a proxy, not the topology of the actual Unknown region or evidence that fine features are safe. Reverse distance and original STEP-to-reference error remain unbounded by this audit.

The new implementation passes 57 Python method tests and 24 proof-auditor tests; 29 native-interface checks are recorded separately. The proof-log audit verifies 36 records and 12 unique partition proofs, with 1,122,136 exact component comparisons. The retained Stage10 audit verifies 27 records, 9 unique proofs and 769,455 component comparisons while preserving its failed integration gate. Fresh same-host

Table S6: Stage11 control runs: all leaves pass in every repeat (no Unknown). Counts are identical across repeats; times are medians of three full CLI walls.

Control	Method	Global queries	Local tests	CLI (s)
Thin slab	E: centre	17,428	17,428	2.084
Thin slab	F: face + reset	16,984	373,480	2.260
Thin slab	W: + walk	15,007	946,151	2.187
Thin slab	MW: + four pieces	15,159	2,105,535	2.232
Separated slabs	E: centre	50,762	50,762	5.135
Separated slabs	F: face + reset	33,930	864,900	4.680
Separated slabs	W: + walk	29,981	1,981,360	4.546
Separated slabs	MW: + four pieces	29,976	4,532,050	4.576

Table S7: P02 cost and partition detail. Search includes inline exact checks. CLI ranges are minima–maxima, not confidence intervals. Peak sum is the largest reported sum of separate self/native peaks over three runs, not simultaneous RSS.

Method	Leaves	Unknown	Search (s)	CLI range (s)	Peak sum (MiB)
E	278,734	85,906	16.11	23.43–24.57	788.7
F	275,222	70,656	18.50	25.99–26.43	839.6
W	285,670	34,975	21.89	29.58–30.36	843.9
MW	277,699	33,892	22.67	30.17–30.84	828.9

compilation of W and MW gives byte-identical node arrays for all six case/variant pairs. These are different software checks, not independent human proof review or independent laboratory reproduction. The exact checker is shared between producer and auditor and is covered by its own unit tests; the audit is not a second implementation of the entire nearest-face search.

The environment is macOS 26.6.2 arm64, Python 3.12.14, NumPy 2.3.5, Apple Clang 21, CMake 4.3.2, Eigen 5.0.1, and libigl revision `182e36df247e3bc14ae7d9a5ea8920908fa5cd15`. The source can be rebuilt without CGAL. The current Python peak-memory conversion assumes macOS byte units and must be adapted before Linux memory comparisons. The artifact guide separates read-only audit from experiments that must write into new directories; formal records must not be replaced.

S7 Stored-data correctness obligations

This section supports the corrected mathematical interface without changing the v13 research questions, populations, tables or timing results. It is not an additional engineering experiment. All inequalities below concern the same observed rows and reference correspondence as the main theorem.

S7.1 Field evaluation, normal storage and zero budgets

The value and gradient contracts in the main paper are statements about real fields. A stored evaluation need not equal that real field value. If v_j is the stored evaluation at $\widehat{p}$, an upper bound on $|v_j - \phi_j(\widehat{p})|$ must include both model and evaluation error. The projection budget can then use this total value-error bound plus $|v_j|$ to bound $|\phi_j(\widehat{p})|$. Similarly, the gradient budget in the direction lemma must apply to the actual vector being normalized, or an additional gradient-evaluation defect must be included. An envelope of the analytic perturbation alone does not establish these stored-data bounds.

The direction lemma bounds $\|n^\circ - n\|_2$ for the exact normalization $n^\circ = v/\|v\|_2$. For the stored normal $\widehat{n}$, a valid route is

$$\|\widehat{n} - n\|_2 \le \eta_{\mathrm{dir}}(g) + \delta_{\mathrm{norm}}, \qquad \delta_{\mathrm{norm}} \ge \|\widehat{n} - n^\circ\|_2.$$

A direct reliable upper bound on $\|\widehat{n} - n\|_2$ is equally suitable. Theorem 6 uses that full row-error budget; it does not assume that a floating-point normalization produces an exactly unit real vector.

Zero perturbation amplitude removes model error, not necessarily arithmetic error. For the unit-sphere SDF evaluated at the off-surface query point $(1, 1, 0)$, the exact gradient, also the normal at its primitive projection, contains $1/\sqrt{2}$ and is not representable by a finite binary fraction. A stored binary64 vector cannot have zero error relative to that normal. By contrast, an exactly represented axis-plane case can have genuinely zero error. Thus neither a blanket zero budget nor an arbitrary positive epsilon floor is justified: an enclosure or an exact equality argument is required.

S7.2 Stored rows, residual and output

For each actual stored offset, define the nonnegative defect

$$\omega_{h,i} = \left| \widehat{h}_i - \widehat{n}_i^{\mathsf{T}} (\widehat{p}_i - c) \right|.$$

With unit reference normal, the main paper's centered-row expansion gives

$$|\widehat{h}_i - h_i| \le \eta_{p,i} + R_i \eta_{n,i} + \omega_{h,i}, \qquad R_i \ge \|\widehat{p}_i - c\|_2.$$

All subtractions and dot products in these definitions are exact real operations on the stored entries. A cell half-diagonal is sufficient only when its relation to the actual stored center and point is established.

For the same stored iterate used to generate the output, set

$$\widehat{H} = \widehat{N}^{\mathsf{T}}\widehat{N} + \lambda I, \quad \widehat{q} = \widehat{N}^{\mathsf{T}}\widehat{h}, \quad \rho = \widehat{H}\widehat{y} - \widehat{q}.$$

The products defining $\widehat{H}, \widehat{q}$ are exact real products of stored rows. Enclosing $\|\rho\|_2$ therefore accounts for the effect of rounded assembly and an inexact solve. A small residual against only the rounded assembled matrix is not a substitute. With $\Delta N = N - \widehat{N}$, $\Delta h = h - \widehat{h}$ and $\widehat{r} = \widehat{N}\widehat{y} - \widehat{h}$, the exact identity is

$$H(\widehat{y} - y) = \rho + \Delta N^{\mathsf{T}}\widehat{r} + N^{\mathsf{T}}(\Delta N\widehat{y} - \Delta h).$$

Only the reference rows are unit; the posterior comparison uses $\|\widehat{N}\|_2 \le \sqrt{m} + \eta_N$, not $\sqrt{m}$. Independently enclosed matrix and posterior bounds may have a different numerical ordering from their exact formulas. Their minimum is valid only when each endpoint bounds the same displacement.

For the actual returned vertex z, a reliable implementation also needs

$$d_{\text{out}} \ge \|z - P_C(c + \widehat{y})\|_2, \qquad \|z - P_C(c + y)\|_2 \le d_{\text{out}} + B_{\text{QEF}}.$$

The iterate, center, row correspondence, regularization and closed box must be identical in the certificate and the actual output. An exact rational ℓ_1 bound on the output defect is a possible conservative route. The target is the box projection of an unconstrained reference QEF minimizer, not a box-constrained QEF optimum or a closest surface point. Admission compares an upper-enclosed error with an exact or lower-enclosed tolerance, including its cell-scale product. No audit slack is added.

S7.3 Upstream prerequisites and evidence boundary

Every selected-path comparison must have a reliable strict margin in the required direction; the projection budget must be strictly below a valid reach bound, and all queried points and boxes must lie in the domains of the corresponding certificates. Computed corner signs must also be established, rather than assumed from an ordinary rounded sign test. Unresolved comparisons, nonfinite input data, unseparated denominators, unsupported domains or a missing output correspondence give UNKNOWN. An unavailable optional bound is distinguished from invalid input and does not remove another independently established usable bound, as detailed below. The strict comparisons at the projected point and continuity establish local boundary visibility even if the projection is on the witness-ball boundary or the ball has radius zero; whole-mesh topology does not follow.

A separate post-v13 implementation constructs pointwise enclosures from known exact analytic reference formulas, using rational endpoints and integer-square-root bounds. Its supported represented inputs are spheres, axis-aligned planes and infinite cylinders, and ring tori about the z axis, with valid radii and tube conditions, under signed hard min/max composition. The strict unit-direction requirement on represented plane/cylinder parameters is not a floating-point tolerance; general rotated directions are not silently normalized and are normally rejected. This is a different, restricted information model, not validation of the historical cell-wide perturbation declarations or an upstream producer for arbitrary sampled or neural fields. Represented-input guarantees also exclude measurement and pre-import quantization error. Local enclosure does not constitute formal verification of the language runtime, extraction pipeline or completed mesh.

The original supporting implementation and its existing records remain identified in the historical dependency manifest. The hardening addendum below has its own source snapshot and software-check records; it does not overwrite that implementation or its saved results. Sections S1–S6 and the historical main results retain their original evidence and limitations. Section S8 is a separate downstream comparison on saved inputs, not a validation of the historical upstream declarations. No new geometry or performance experiment, independent human proof review, or whole-pipeline formal verification is added here.

S7.4 Finite endpoints and exception-state recovery

The hardening addendum separates three events: computing an enclosure, obtaining a finite usable endpoint, and satisfying the complete admission test. A positive Boolean field describing the first event is insufficient for either of the others. At the budget interface, the consumer checks the actual matrix, posterior and minimum endpoints against their status records. The exported minimum must agree with the minimum of the two exported route endpoints. These routes enclose the same displacement, with the same stored rows, iterate, center and regularization.

An exported $+\infty$ route is marked unavailable. If the other route has a valid finite endpoint, it remains usable; if neither does, admission is blocked. NaN, negative bounds, $-\infty$, missing required fields and

inconsistent records are invalid, not alternative finite certificates. In particular, a completed rational calculation whose minimum overflows binary64 retains its computation record but has no finite usable bound. The compatibility verification flag is then false. A finite bound still requires the geometric premises, the actual-output transfer budget and the exact or lower-enclosed tolerance comparison. This interface check assumes a trusted budget producer: it is not an independent proof of an arbitrarily forged finite endpoint or a data-authentication mechanism.

Every repeated admission attempt first invalidates previous decision flags, budgets, output bindings and diagnostics. Expected input/type and arithmetic failures clear partial evidence and record their type and reason as an UNKNOWN outcome. Unexpected programming errors and interrupts also clear the state but are re-raised, rather than silently treated as successful completion. Unestablished row budgets are absent, not zero. Old oracle diagnostics are removed on reevaluation and never supply admission premises. The raw constraints and selected-leaf identity remain available for a fresh binding check. A center fallback is an uncertified display vertex only. This policy concerns an admission attempt on a well-formed cell record; it is not a recovery protocol for arbitrary malformed objects or concurrent mutation.

S7.5 Bounded regression evidence for the hardening

The separate addendum changes two runtime modules. It passes 140 test methods: 126 retained tests and 14 new methods with multiple subcases. The 14 new methods also pass with Python assertions disabled; these repeated checks are not another independent test population. They exercise endpoint overflow, NaN and negative-bound rejection, inconsistent flags, and a legitimately unavailable route paired with a usable finite route. Fault injection covers five expected exception classes at six call stages, including output transfer, as well as failure after candidate/budget assignment, unexpected errors, interrupts and reuse of the same cell after recovery. Expected failures leave no previous or partial admission evidence in the tested record fields; unexpected failures remain visible to callers.

Two existing one-cell software fixtures, a plane and a tiny sphere, were also compared before and after the change. Their selected vertex, admission, row and budget records agree exactly under the comparison script; both one-cell fixtures have empty face arrays, so this is not a mesh-topology test. This is not a replay of the 200-fixture study or the historical CAD cohorts. The source-protection check found the original v15 file set and all 86 file hashes unchanged. The addendum contains the patch, tests and verification receipt. These bounded regressions support the stated interface behavior; they neither exhaust all failures nor prove the language runtime, upstream geometry or the complete extraction pipeline correct. Historical admission, coverage and timing counts are unchanged in this revision.

S7.6 Version-gated saved-record repair and its limits

The artifact retains distinct historical and reliable-data meanings. The v13.3.4 addendum repairs their import boundary in an independent runtime; it does not overwrite the frozen code or retrospectively change its evidence.

Route	Required interpretation	Guarantee boundary
Legacy saved-cell replay	Explicit legacy mode for untagged historical records; original normalization, offset and budget conventions.	Historical replay only. Reliable fields mixed into this format are rejected, not silently ignored.
Reliable saved-QEF replay	Versioned actual rows; recomputed residual, row-offset corrections, QEF bound and candidate-output transfer.	Conditional candidate bound. Neither source authentication nor renewed verification of the geometric error budgets.
Immutable downstream API	Validated rows, offsets, budgets, iterate and λ; explicit unit-reference-row premise.	Declared unconstrained-QEF displacement, not a mesh-vertex admission.
Reliable online admission	Supported analytic producer and all geometric, numerical and actual-output guards.	Candidate binding is distinct from final returned-vertex certification.
Common-box pilot	Source-specific rows frozen once for the same component boxes.	Unchanged S8 comparison, not historical upstream recertification.

Defects retained as regression cases. The earlier policy adapter treated the nonempty string `"false"` as true. The old cell validator also ignored an explicit observed-row field while reconstructing normalized constraints. A mixed-format record could therefore receive a bound for another system. That record violated the theorem's correspondence and budget premises; it was an interface counterexample, not a theorem counterexample. Conversely, a legitimate reliable-producer record failed the old offset-budget

check. The frozen old code retains these behaviors for provenance; the new public facade does not infer that an incomplete new record is a valid old one.

Implemented boundary checks. The legacy format is `qef-cell/legacy-v9`; the reliable format is `qef-cell/reliable-v1`. Both have explicit row-semantics labels. An untagged record requires an explicit legacy argument. Unknown versions, mixed fields and missing reliable rows are rejected. Policy guards must be actual Booleans; only true premises allow subsequent admission. Unbound cached bound dictionaries are refused by the repaired old-policy entry point. The immutable API's input/program-keyed stage cache is separate.

Reliable exports carry a canonical digest of the represented rows, offsets, budgets, iterate, λ, center, box, declared correspondence and output identity. Replay checks this digest, rechecks the rounded assembly diagnostics, and recomputes the exact stored-row residual and reliable closure. Stored residuals and pass flags alone are not sufficient. A changed row, budget, iterate or λ cannot retain its old numerical certificate just by recomputing the digest. Candidate transfer is recomputed against the real box projection, with a separately checked output budget.

The export distinguishes `candidate_output_binding_verified` from `returned_vertex_certified`. The old binding name is retained only as a candidate-scoped alias. If completion returns a fallback vertex, the returned-vertex certification is false, even when the rejected candidate still has a valid numerical displacement bound. Offline `Verified` means a conditional bound for that candidate; it is not a new geometric admission and does not certify the fallback vertex.

Bounded regression and compatibility evidence. The repair passes 171 test methods: the retained 140 and 31 new interface methods, with multiple malformed-input and target-change subcases. The 31 new methods also pass with assertions disabled; this is a repeat check, not another independent population. A fresh type scan again finds proper Booleans for all 172,828 premise flags in 43,207 historical records.

For 96 existing historical records, old/new replay preserves the normal rows supplied to the analytic policy, numerical bounds and three-threshold decisions. The primary historical normalization and the legacy interval replay use their own original operation orders; replacing one by the other can change final floating-point bits. They are retained separately, not used to reconstruct new reliable records. The initial attempted shared normalization and its detected numerical differences are retained in the repair log.

An explicit compatibility exercise upgrades copies of the identified v15 fixture source, adding version/status metadata without changing its rows or candidate budgets. All 200 numerical replays match those budgets; 158 source outputs remain recorded as certified and 42 remain uncertified. This is not automatic migration of arbitrary input, a new geometric extraction, or evidence that all 200 outputs meet an admission tolerance. Neither these replays nor the earlier 296-record S8 audit reruns the full historical CAD cohorts, field-budget certification or timing experiment.

Residual trust boundary. A digest detects stale or mismatched contents; it is not a signature or a proof of geometric provenance. A party that regenerates a complete record can also regenerate its digest. The replay therefore continues to assume that the declared point/normal budgets cover the stated reference rows. Independent upstream authentication, arbitrary sampled/neural fields, concurrent mutation, global mesh validity and full program verification remain outside this repair. No theorem, historical scientific table or timing claim is changed by these interface regressions.

S8 Matched common-box numerical pilot

This new analysis keeps its inputs, protocol and results separate from S1–S7. The protocol was locked locally before the new comparator was evaluated, not publicly preregistered. Historical eligible identifiers are sorted by integer cell tuple; task indices are $\lfloor k(n-1)/7 \rfloor$, $k = 0, \ldots, 7$, with duplicates removed for short lists. Twelve tasks contribute eight records each; M05 contributes none. The frozen file contains full observed rows, declared radii, candidate, center, iterate and box, but no oracle error. The 200 saved resolution-6 v15 cells are a distinct software-fixture stratum, with all unresolved records retained. No scientific extraction is rerun.

Row reconstruction is source-specific. Historical v9 assembly normalized each constraint normal before forming its offset and Gram matrix; the historical stratum reconstructs those assembly rows with the same legacy operation. The v15 fixture stratum takes the explicit stored observed rows without renormalizing them. These rows are then frozen, and all three controls consume the identical frozen entries. The comparator does not invoke the legacy saved-cell validator discussed in Section S7.6.

S8.1 Common information and exact certificate

Put $N = N_0 + D$, $h = h_0 + d$, with $|D_{ij}| \le e_i$ and $|d_i| \le f_i$. All stored numbers mean their exact represented values; $\lambda > 0$ is common. The general-row analytic control uses $a_i = \sqrt{3}e_i$, $a_N = \|a\|$, $\beta = \sqrt{\sum_i (\|n_{0,i}\| + a_i)^2}$ and $E_r = \|(a_i\|y_0\| + f_i)_i\|$. Its posterior numerator is

$$\|H_0 y_0 - q_0\| + a_N \|N_0 y_0 - h_0\| + \beta E_r.$$

The matrix numerator is $\|H_0 y_0 - q_0\| + (2\beta a_N + a_N^2)\|y_0\| + a_N\|h_0\| + \beta\|f\|$. Their minimum divided by λ is valid for this common-box family. The original reference-unit assumption is not used on only one side.

Set $R = H_0^{-1}$ using exact rational elimination and $e = y - y_0$. Then

$$e = (I - RH)e + R(q - Hy_0).$$

Natural interval assembly gives nonnegative majorants M_0, b_0. With

$$k = \max_i \sum_j (M_0)_{ij} < 1, \qquad B_{\text{scalar}} = \frac{\sqrt{3}\max_i (b_0)_i}{1 - k},$$

the scalar control returns B_{scalar}. The componentwise control also encloses the identities

$$\begin{aligned} I - RH &= I - RH_0 - \sum_i \big[(Rn_{0,i})D_i^{\mathsf{T}} + (RD_i)n_{0,i}^{\mathsf{T}} + (RD_i)D_i^{\mathsf{T}}\big], \\ R(q - Hy_0) &= R(q_0 - H_0 y_0) \\ &\quad + \sum_i \big[(Rn_{0,i})(d_i - D_i^{\mathsf{T}} y_0) + (RD_i)(h_{0,i} - n_{0,i}^{\mathsf{T}} y_0) + (RD_i)(d_i - D_i^{\mathsf{T}} y_0)\big]. \end{aligned}$$

Taking componentwise minima of the two valid absolute majorants yields M, b for the same family. Rational solves propose $w = (I - M)^{-1}\mathbf{1}$ and $u = (I - M)^{-1}b$, but the final predicate separately checks $M, b, u \ge 0$, $w > 0$, $Mw < w$ and $b + Mu \le u$. The positive weighted contraction implies $\rho(M) < 1$ and $(I - M)^{-1} = \sum_{k \ge 0} M^k \ge 0$. Applying this inverse to $|e| \le M|e| + b$ proves $|e| \le u$. Singular solves or failed checks return Unknown; no arbitrary epsilon is added.

Algebra is rational, square-root brackets are integer-checked dyadic enclosures, and conversion to a binary64 upper endpoint is checked against the rational endpoint. A zero input radius retains the exact stored-system residual. For admission, an exact ℓ_1 upper bound on the difference between the actual candidate and exact clamping of $c + y_0$ is added. The three thresholds are exactly $0.1s_C$, $0.25s_C$ and $0.5s_C$, where s_C denotes the largest exact stored box width. These are common diagnostic thresholds, not bit-identical historical floating-point threshold products.

S8.2 Separate fixture results and audit scope

For the 200 fixtures the componentwise bound is smaller than both controls in every case; median componentwise/analytic and componentwise/scalar ratios are 0.240 and 0.289. The 164 admissions at $0.5s_C$ are conditional on the 164 recorded non-QEF prerequisite passes. They do not replace a previous runtime count under different budgets or tolerance construction.

The artifact retains 12 focused tests, replay of every method output, 292 exact componentwise certificate checks, and independent rational Cramer-rule solutions at four fixed box corners per input. All 1,184 analytic, 1,056 scalar and 1,168 componentwise finite-bound probes pass. Those corners need not be unit normals: they are probes in the deliberately shared box family. They are neither exhaustive uncertainty tests nor extra CAD observations. The universal implication is the argument above, not the finite probe count. The exact arithmetic implementation is part of the trusted software base.

Table S8: Separate downstream control on all 200 existing small-grid software fixtures. Only 164 have the required non-QEF prerequisites recorded; the other 36 remain Unknown for admission even with finite downstream bounds. These are not 200 CAD models or a rerun of upstream certification.

Method	Finite	No finite bound	$0.1s_C$	$0.25s_C$	$0.5s_C$
Analytic common-box	200	0	10	144	155
Scalar contraction	200	0	70	138	154
Componentwise	200	0	149	164	164

Saved historical oracle displacements are consulted only after decisions for a floating-point diagnostic. They do not establish upstream budget validity or replace exact geometry. Only aggregate reproduction wall time is logged; there is no formal timing or production-performance claim. The 96-cell pilot must not be pooled with the 200 fixtures or extrapolated to all historical cells. Rump/INTLAB implementations, correlation-preserving alternatives and a held-out geometric evaluation remain distinct follow-ups.

S9 Engineering readiness and locked recomputation protocol

The new engineering action is a source/interface preflight, not a completed utility experiment. It reads 31 scoped files from the retained engineering stages, corrected arithmetic interfaces and fixed mesh inputs. The inputs exist, but six required interfaces are not established: row-to-final-vertex identity; a supported reliable field/normal producer for P02; a valid reference-fit budget; compatible local recomputation; a complete two-way surface checker; and a shared full-cost ledger. The preflight therefore returns BLOCKED, with zero utility runs. Fifteen software tests check this fail-closed behavior, not engineering accuracy or speed.

The mismatch is substantive. Some earlier repaired surfaces use root-point piecewise-linear output or tetrahedral contraction, whereas the corrected QEF theorem compares corresponding regularized systems with common-box projection. P02's mesh-derived distance and pseudonormals do not become supported analytic SDF declarations merely because the model imports. Stage11 verifies only one mesh-distance direction and still has Unknown patches. These components cannot be joined without checking their meanings.

S9.1 Fixed-topology geometric bridge and cost obligation

Let M_* and M share all vertex identifiers and triangle connectivity. If every actual vertex displacement is bounded by B_v, a pair of points with the same barycentric coordinates on corresponding triangles differs by at most $\max_v B_v$. In both directions this proves

$$d_H(M, M_*) \le \max_v B_v.$$

Only with an additional valid $d_H(M_*, S) \le \delta_{\mathrm{ref}}$ can one infer $d_H(M, S) \le \delta_{\mathrm{ref}} + \max_v B_v$. All auxiliary and fallback vertices must be covered. Changed connectivity or independent refinement requires a different correspondence argument. The preflight does not construct M_* or δ_{ref}. If doing so requires expensive reference queries, those are charged, not counted as work avoided later.

S9.2 Protocol retained without a positive outcome

The locked candidate workload contains the saved two slab controls and P02, with exact represented $h/2$ as the primary threshold and $h/4, h$ as fixed sensitivities. The intended target in this protocol is the stored reference triangle mesh, not the original STEP/B-rep. Uniform expensive recomputation, QEF-guided recomputation and a cheap residual or direct-witness control must use the same geometric accuracy target and final checker. A heuristic that fails the final checker must repair its output and pay that cost. No method receives free offline oracle labels. Producer, verification, recomputation, fallback, final checking and I/O are separately counted and included in the declared total; no artificial delays represent expensive work.

The protocol and executable input guard are packaged, but no full geometry runner is claimed. Missing evidence leaves the run unstarted; a boolean assertion of readiness is not accepted as the missing producer or proof. An initially proposed independent analytic patch pilot was withdrawn before execution because it did not test this QEF bridge. It has no results and is excluded from the paper's evidence. The protocol makes the missing engineering interface explicit and auditable; it does not close the engineering validation requirement.