Design and evidence-gated refinement of a 3-D-printable smooth-tip toroidal rotor

The Poltergeist rotor program: a reproducible diagnostic improvement, its costs, and the evidence still missing

Status: research report built from the evidence-audited kit in `docs/rotor-paper-kit/`
Written at: branch `rotor-research-report`, based on the canonical rotor history consolidated at `b508f68` and extended through the qualification disposition (`c881754`) and the rejected steady-pilot record (`5a7e337`) Evidence manifest: `sim/toroidal-research/paper-v1/data.json`, regenerated and drift-checked with `npm run aero:paper-data -- --verify`

Abstract

This report documents the design and computational refinement of an 88.9 mm (3.5 in) smooth-tip toroidal rotor for a sub-250 g quadrotor at a frozen comparison point of 24,486 RPM. The product hypothesis was that a closed, smoothly turned tip could improve contact robustness and reduce tip-noise mechanisms while retaining useful aerodynamic performance. Development began with parametric geometry, manufacturing, and structural surrogates, but early review exposed a more basic gap: the rotor had not demonstrated credible lift. The program therefore built an auditable, headless evidence ladder combining a frozen commercial control, signed low-order loading, low-Reynolds-number section analysis, targeted SU2 experiments, XFOIL correlation, and exact-geometry DUST unsteady vortex-lattice/free-wake simulation. A preregistered 3,600-cell sectional-CFD campaign proved computationally and numerically unproductive — its two pilots admitted zero coefficients and its measured wall times imply 12.6–26.1 serial days — and the program was redirected to bounded geometry batches, causal diagnostics, and mature-wake promotion rules. A repeatable 4.089 N cold-start result, briefly 143% of the commercial thrust bar, was rejected after a wake-duration ladder decayed to 1.178 N at two revolutions. Subsequent broad, factorial, airfoil, and rotor-level batches produced a retained 6%-camber, 7-inch-pitch, –8-degree-incidence-split closed-tip rotor. At revolution four it produced 2.441372 N and 94.032 W of resolved shaft power in the frozen DUST workflow: 6.37% more thrust than the prior TR-05M leader, but 14.92% more power, 7.44% lower thrust per shaft watt, and only 85.32% of the measured HQProp thrust target. The mature-wake gate remained closed. Two subsequent frozen batches returned zero leader improvement, firing the preregistered diminishing-return stop rule. The result is therefore a reproducible design hypothesis and a methodological advance, not evidence of superior efficiency, noise, strength, or flightworthiness. The next decisive work is independent viscous validation, proof-spin and structural testing, and matched-RPM thrust, torque, power, and acoustic measurement.

Keywords: toroidal propeller; closed-tip rotor; low Reynolds number; free-wake vortex lattice; DUST; XFOIL; SU2; preregistration; evidence-gated optimization; negative results; additive manufacturing; small UAS

Governing thesis. We began with a compelling product constraint and a plausible shape, learned that confidence had been built in the wrong order, constructed an evidence system strict enough to reject our own apparent breakthrough, redirected a stalled program toward bounded exact-prop experiments, obtained a reproducible but qualified improvement, and stopped when the evidence said local tuning had run out of value.

How to read this report

Every number in this report is extracted from a frozen, committed artifact — an immutable ledger checkpoint, a frozen protocol or report JSON, or the registry — through a deterministic extractor (`scripts/build-rotor-paper-data.mjs`) that hashes its 24 sources, asserts identity and denominator invariants, and fails on drift. The extractor's output manifest and the ten generated figures live in `sim/toroidal-research/paper-v1/`. Task-history transcripts (exported in `docs/rotor-paper-kit/transcripts/`) are used for decision rationale and chronology only, never for numbers; where a live transcript statement and a committed report disagree, the report wins and the discrepancy is stated.

Three vocabulary rules apply throughout. DUST results are *unsteady vortex-lattice/free-wake* results — an inviscid model, never "CFD" without qualification. "Resolved shaft power" is torque-derived power inside that inviscid model, not electrical input or a bench measurement. And the retained rotor is a *diagnostic leader* — an exploratory, non-admissible simulation result — not an optimized, validated, or flight-ready propeller.

---

1. Introduction

Poltergeist is a parametric design studio for sub-250 g 3-D-printable FPV quadrotors. In August 2026 it grew a rotor program with an unusual premise: the propeller topology was a product constraint before it was an aerodynamic conclusion. The user wanted a small printed rotor whose blade tips close into smooth loops — fewer sharp exposed ends around people and obstacles, greater geometric continuity and potential stiffness, and a plausible (hypothesized, not demonstrated) path to lower tip-vortex noise — integrated with the aircraft configurator rather than imported as a foreign STL.

That framing fixes the research question. It was never "what unconstrained propeller makes the most thrust?" It was: *can this closed-tip family become aerodynamically credible, and can it approach or beat a good store-bought propeller without surrendering the reasons it was chosen?* Conventional open-tip propellers therefore appear throughout this report as controls — they are mandatory comparisons, not eligible substitutes.

Concretely, the program asked four questions at a frozen comparison point of 88.9 mm diameter and 24,486 RPM:

1. Can the exact closed-tip geometry produce credible thrust at all?
2. Which geometry and section variables improve mature-wake loading?
3. Can an improvement survive power, wake, manufacturing, structural, and evidence gates simultaneously?
4. Does the result beat the measured commercial control?

The answers, in order: yes, with heavy qualification (§6.5–§6.12); global pitch, front/rear incidence split, and lifting-section camber, at a disproportionate power cost (§6.10); no — the full mature-wake gate never opened and no physical gate was attempted (§6.12); and no — the retained design reaches 85.32% of the measured control thrust (§6.12).

What makes the program worth reporting is not the headline thrust number. It is the order in which confidence was built, torn down, and rebuilt. The program produced a visually compelling parametric rotor before it had lift evidence; corrected into a validation campaign so rigorous it stopped optimizing the propeller; produced a repeatable simulation result 43% above the commercial bar and then rejected it as a cold-start transient; replaced an unbounded search with preregistered factorial batches and frozen promotion rules; obtained one real mature-wake improvement; and stopped by rule when two consecutive batches returned nothing. The methodology — preregistration, fail-closed admission, immutable ledgers, exact cryptographic identities, explicit failure populations, short-screen versus long-wake promotion, and a diminishing-return stop rule — is presented as a first-class finding (§4, §8.5), because it is the part of the work most transferable to other small-team, agent-assisted engineering programs.

The report proceeds as follows. §2 places the work against published toroidal, low-Reynolds-number, and small-rotor literature. §3 states the product constraints and design space. §4 describes the methods actually executed. §5 reconstructs the program chronologically, including the human interventions and agent handoffs that changed its course. §6 presents results in the order the evidence matured. §7 catalogs failures and course corrections with equal status. §8–§10 discuss interpretation, limitations, and future work. Appendices give the retained candidate's exact identity, reproduction commands, and the task/commit record.

---

2. Related work

2.1 Toroidal and joined-tip propellers

The most widely known closed-tip propeller work is MIT Lincoln Laboratory's: US Patent 10,836,466 and an accompanying technology summary. Both must be cited for what they are. The patent reports closed toroidal blade forms, design variables (element count, airfoil, sweep, twist, material, diameter), claimed stiffness and acoustic mechanisms, and *selected examples* with favorable thrust-versus-sound relationships. The technology summary is product-level framing. Neither is a peer-reviewed performance paper, neither establishes a universal toroidal efficiency advantage, and the 40–60% figure that circulates in secondary coverage concerns a ring-wing lift/drag example — it is not a propeller-efficiency improvement and is not claimed here. This correction mattered inside the program: the literature spike (§5.8) went looking for "the MIT toroidal paper" and found that the primary public record is the patent and summary.

Independent literature supports both sides of the closed-tip hypothesis. Wei et al. (Drones, 2024) report a ~208 mm toroidal rotor with similar figure of merit and reduced radial/axial sound pressure under equal-thrust conditions — evidence that spectra and directivity, not one scalar, are the right acoustic currency. An experimental counterexample (Mariuta, Ignat, and Cican, 2026) shows toroidal topology does not guarantee efficiency: larger wetted area and joined-blade interaction impose real costs. Boxprop research (Chalmers) demonstrates that a joined tip can remove the conventional open-tip vortex while still producing blade-passage interference, unloading, and swirl. Chen et al. (2025) report a ~254 mm toroidal configuration in which the leading blade substantially reduces trailing-blade lift — directly foreshadowing the rear-system loss this program measured in its own wake diagnostics (§6.6). Larger-scale design-variable studies (Kim et al. 2025; Lee and Choi 2025; Ali et al. 2026) treat pitch/twist, blade angle, loop gap, and front/rear spacing as coupled causal variables, but their optima do not transfer across scale.

2.2 Low-Reynolds-number section behavior

At this rotor's scale, lifting sections operate near Reynolds 20,000–40,000, where familiar airfoil intuition degrades. UIUC small-propeller measurements (Deters, Ananda, and Selig, 2014) show small propellers routinely below Re 100,000 with laminar separation and blade deformation materially affecting performance. NASA's very-low-Re review (Koning, Romander, and Johnson) documents rapid degradation of smooth conventional sections below Re 100,000 and non-reattaching separated shear layers below roughly Re 50,000. Wind-tunnel comparisons (Traub and Coffman, 2019) show thin plates and circular arcs can outperform thick sections at Re 40k–80k, with XFOIL agreement dependent on geometry and separation state. Klose, Spedding, and Jacobs show that at Re 20,000 small incidence changes can reorganize separation and wake topology entirely. Together these set the epistemic budget for §4.3: XFOIL is a bounded screening tool here, not truth, and no smooth polar can be trusted across a state change.

2.3 Wake maturation, hover loading, and acoustics

Optimum-hover loading theory (NASA TM-20250009190) supplies twist/loading priors for lifting legs — a prior, not a validation of any particular pitch law. Propeller free-wake practice (NASA aero-propulsive comparison studies, preserved in the wake-convergence task, §5.11) commonly runs several spin-up revolutions plus averaging revolutions, judging convergence on revolution-averaged and phase-resolved statistics rather than a fixed count; a supplementary research task in this program recommended roughly 6–8 spin-up plus 2–3 averaging revolutions for an ordinary isolated propeller. This context strengthens, rather than excuses, the central caveat of this report: the program's four-revolution promotions are shorter than common practice, and its own wake gate confirms the solution was still changing (§6.12). Small-UAS acoustic literature (NASA measurement, psychoacoustics, and broadband-noise studies) informs the future test design in §10; no acoustic data of any kind exists for this rotor.

2.4 The gap

No published result directly validates an 88.9 mm, printed, three-loop, front/rear-interacting closed-tip rotor at 24,486 RPM. Published toroidal advantages are geometry- and operating-point-specific. That gap is why the program built its own evidence system rather than borrowing conclusions.

---

3. Requirements and design space

3.1 Product constraints

The rotor is a subsystem of a configured aircraft, not a free-standing shape. Binding constraints, inherited from the printed-propulsion epic (`toroidal-rotor-epic.md`):

- closed, smoothly turned tips (the defining product constraint — §5.8 records the explicit decision to keep them when the literature challenged the optimizer's favorite geometry);
- airframe fit: diameter, dynamic growth, deflection, and neighboring-rotor and board clearance coupled to the generated geometry;
- printability as authored: one-piece load paths, minimum station thickness, trailing-edge printability, leading-edge radius, orientation, and process tuple; the audited minimum station thickness of the retained design is 1.284 mm;
- hub continuity and burial without bore or face escape;
- structural plausibility at roughly 114 m/s tip speed, ~30,000 g tip acceleration, 408 Hz shaft rate, and 1.22 kHz blade-passage frequency — numbers that justified balance, fatigue, moisture, resonance, and proof-spin attention from the first week;
- material/process realism: PAHT-CF was recommended as a *development* candidate, explicitly not a flight endorsement, because the early structural model

was a centrifugal surrogate blind to defects, clamp loads, fatigue, creep, moisture, aerodynamic bending, and process variation.

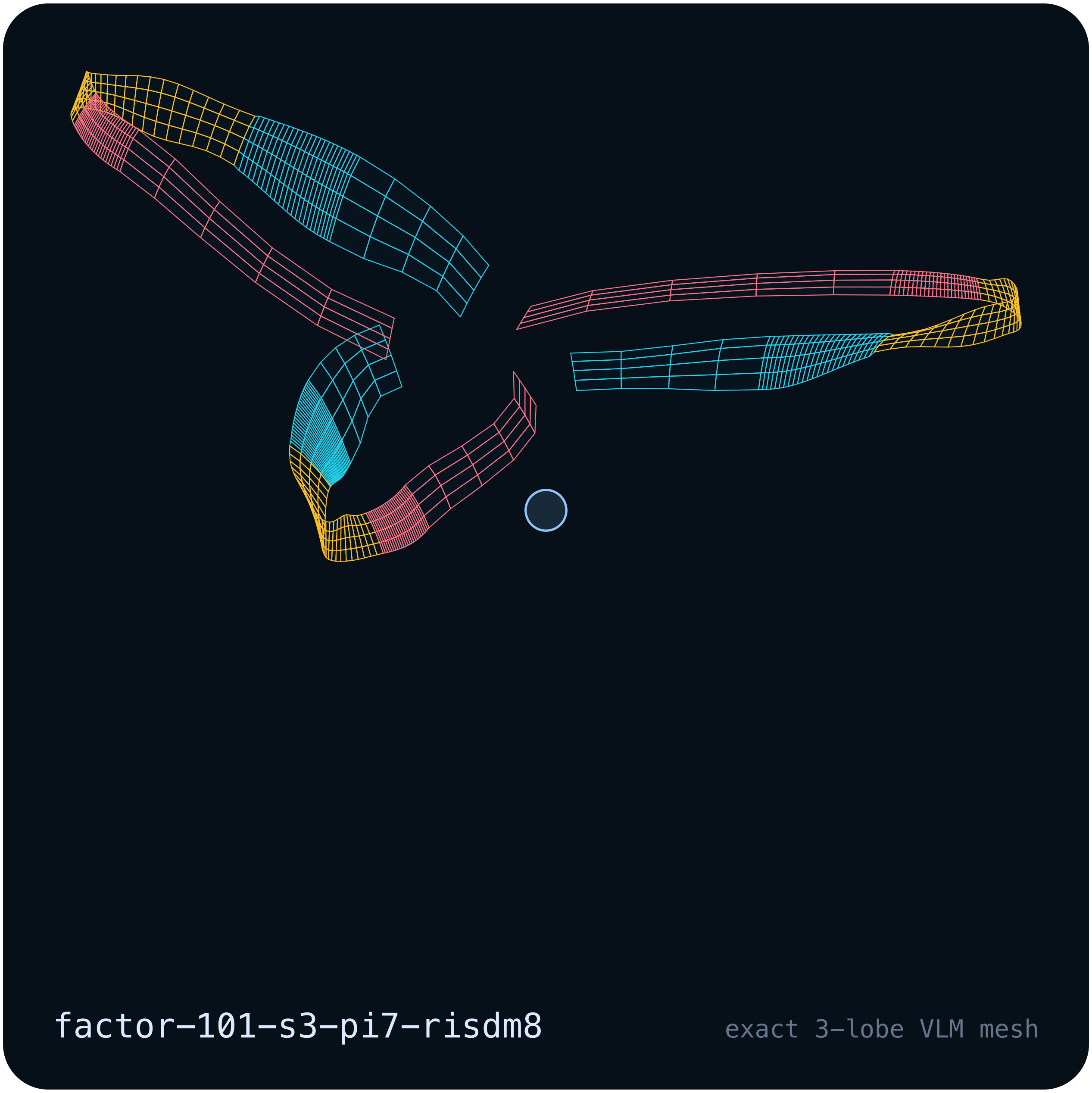
3.2 Aerodynamic design variables

The parametric family exposes: loop count; global pitch; front/rear incidence split; per-region chord scales (front leg, rear leg, root transition, smooth tip-bridge); lifting-section camber, camber position, and thickness; turnover-width scale; turnover-orientation bias; and loop height. The regions have distinct jobs, and the program's design model sharpened into exactly that vocabulary (§5.8): the **front leg** carries the majority of useful loading in the cleanest inflow; the **rear leg** flies in the front leg's downwash and unsteady wake and must be designed independently; the **turnover bridge** is a smooth structural/acoustic connector whose inviscid curved-panel loading must not be "exploited" as if it were validated lift; the **roots** are non-lifting load-transfer regions until three-dimensional evidence supports another treatment.

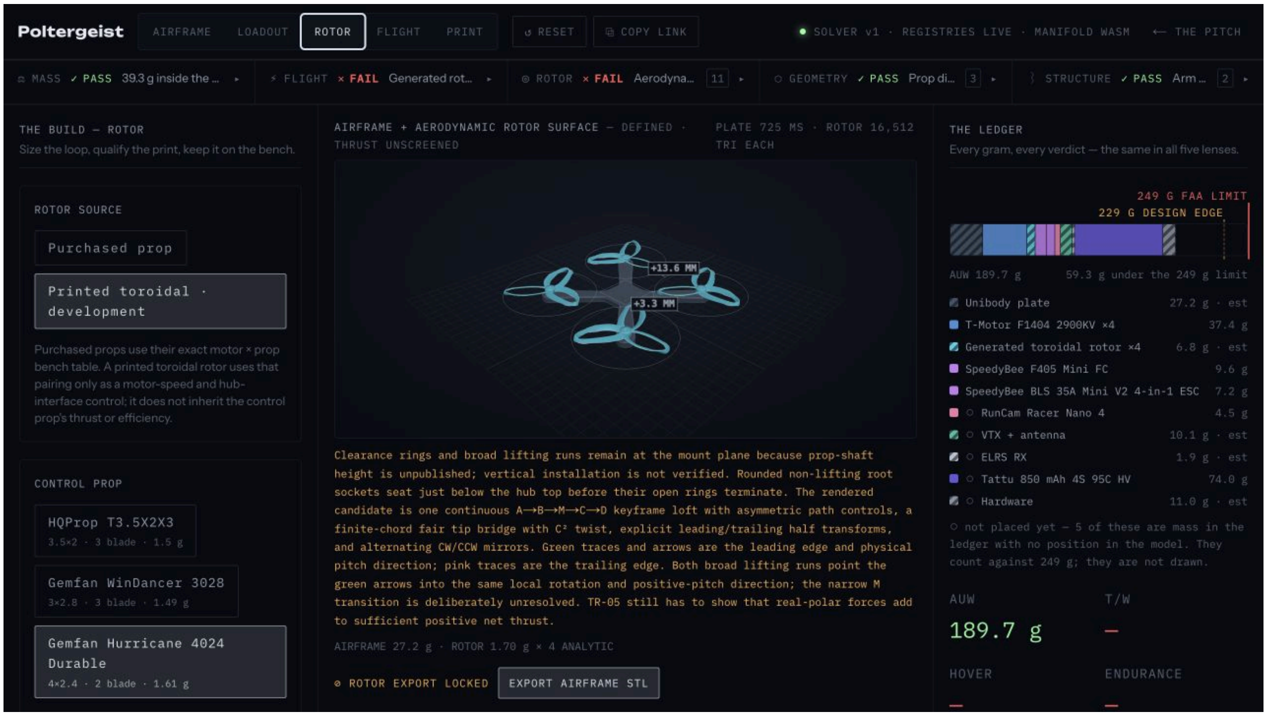
3.3 Identity discipline

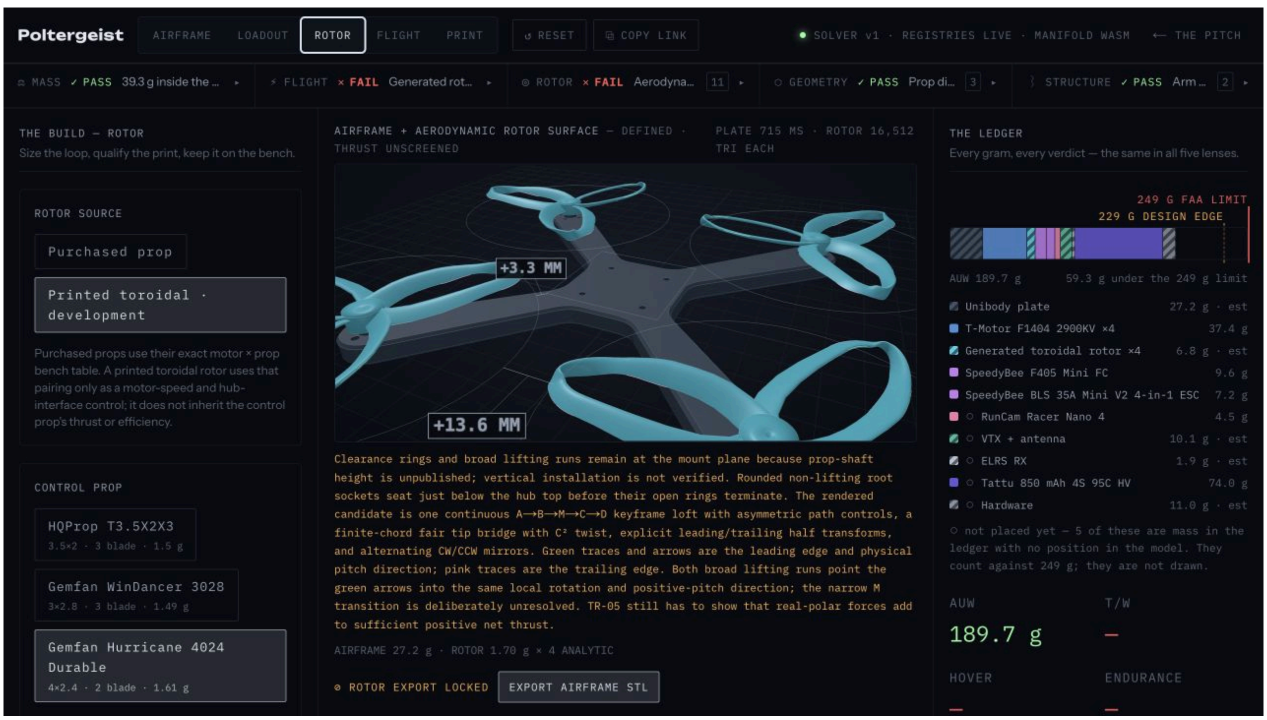
Every exact candidate carries a deterministic geometry identity (SHA-256 over the canonical geometry serialization) and, once meshed for DUST, a mesh identity. Evidence attaches to identities, not names: a visually similar rotor cannot inherit results from another hash, and any geometry change creates a new identity that invalidates downstream evidence. This is what later allowed one physical design to be recognized under three batch aliases (§6.11) and what makes the first-print critical path's candidate freeze meaningful (§6.12). The same discipline exposed a cross-platform mesh-serialization mismatch (§7, F-07) that a looser regime would never have seen.

The retained candidate's exact render:



Earlier geometry stages — useful for the debugging narrative in §5.1, and explicitly *not* renders of the retained design — are preserved in the kit:



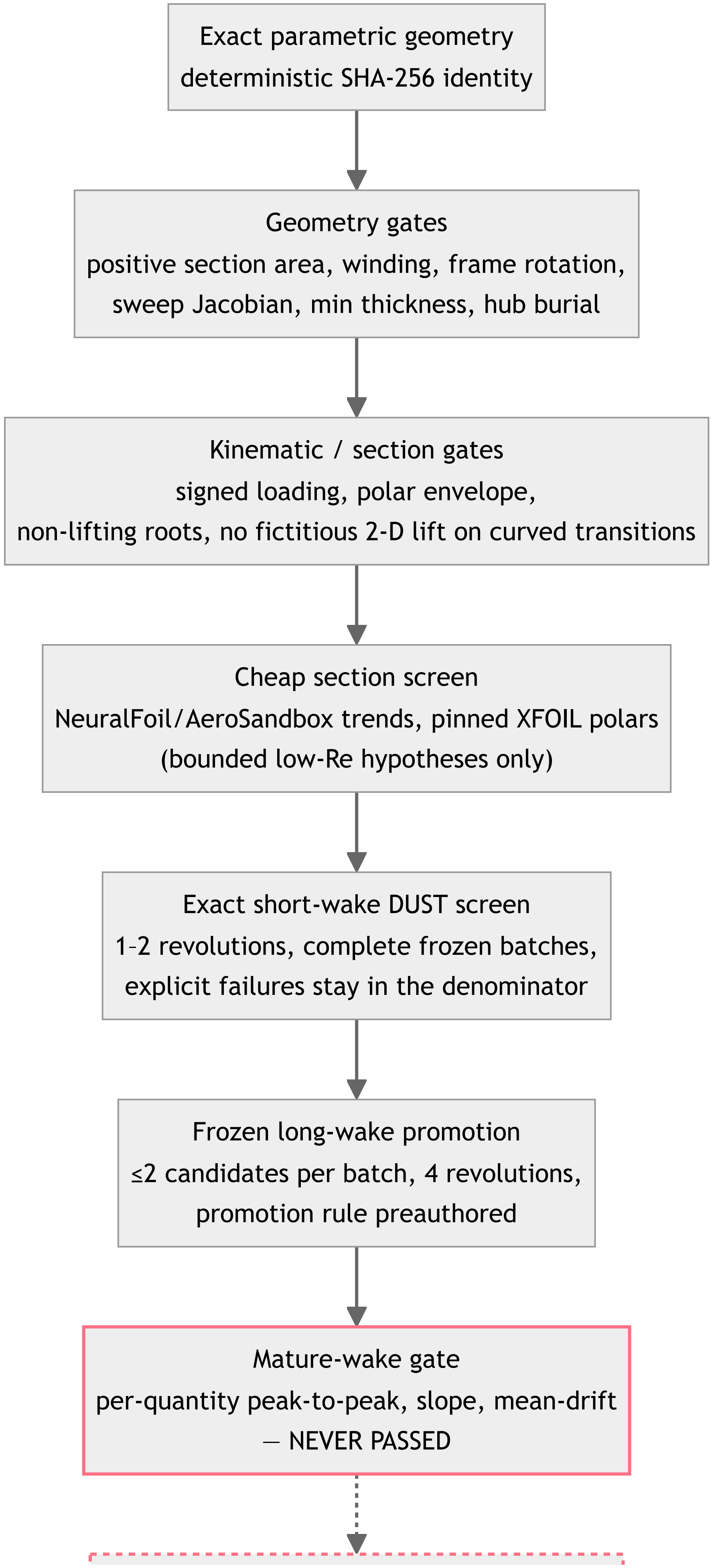


4. Methods

4.1 The evidence ladder

The program's central mechanism is a fail-closed ladder: a candidate earns its way from geometry to qualification, and failure at any rung leaves it — with its failure recorded —

at that rung. Nothing skips ahead, and no downstream gate can be satisfied by upstream enthusiasm.


Exact parametric geometry
deterministic SHA-256 identity
Geometry gates
positive section area, winding, frame rotation,
sweep Jacobian, min thickness, hub burial
Kinematic / section gates
signed loading, polar envelope,
non-lifting roots, no fictitious 2-D lift on curved transitions
Cheap section screen
NeuralFoil/AeroSandbox trends, pinned XFOIL polars
(bounded low-Re hypotheses only)
Exact short-wake DUST screen
1–2 revolutions, complete frozen batches,
explicit failures stay in the denominator
Frozen long-wake promotion
≤2 candidates per batch, 4 revolutions,
promotion rule preauthored
Mature-wake gate
per-quantity peak-to-peak, slope, mean-drift
— NEVER PASSED

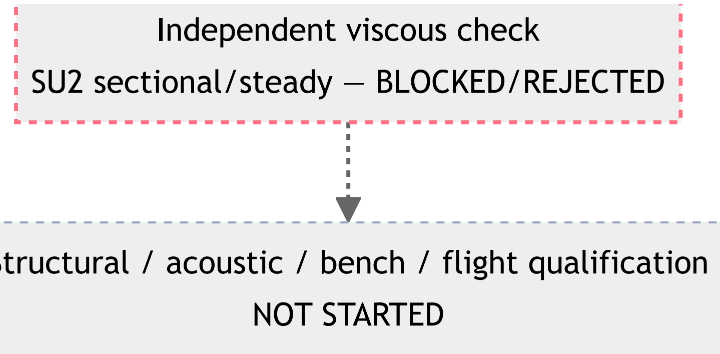


Figure 1 — the evidence ladder. The retained design stops at the long-wake promotion rung: it is the best candidate ever promoted, and it fails the mature-wake gate above it. Everything past that rung is future work.

4.2 Controls and operating point

The commercial control is frozen and singular: an HQProp T3.5x2x3 on a T-Motor F1404 2900KV motor at 16.7 V and 85% throttle, from the manufacturer's bench table — 291.77 g of thrust (2.8613 N) at 6.18 A and 24,486 RPM. The registry row (`registries/motors.json`) is the committed source; the extractor verifies 291.77 g × 9.80665 m/s^2 agrees with the frozen 2.8613 N target to within 5×10^{-4} N.

Two provenance qualifications bound every use of this number. First, the bench table reports **electrical input**, not shaft power, so no efficiency comparison against DUST's resolved (torque-derived, inviscid) shaft power is admissible — the control is a **necessary-condition thrust bar only**: a candidate that cannot approach it in its own favorable model has no case, but matching it in DUST would still prove nothing about the bench. Second, the comparison is matched-diameter and matched-RPM by intent, but motor, voltage, measurement chain, and installation differ; the program never treats the ratio as a same-conditions experiment.

Commercial *geometry* controls (authoritative APC tables; a T-Motor STEP ingestion proof) were built as reproducibility results for the toolchain. They are not evidence that the low-order model successfully predicted commercial static performance — §6.2 shows it did not.

4.3 Models and tools actually used

Tool	Role actually played	Fidelity / limits
TypeScript geometry + signed loading kernel	exact parametric geometry, deterministic identities, signed force integration, envelope gates	kinematics and bookkeeping, not flow physics
XFOIL (pinned, Dockerized)	polar provenance for the low-order model; bounded section correlation (TR-06E); manufacturable-section and local-camber batches (TR-06G/H)	2-D, low-Re-fragile, shape- and state-dependent agreement
NeuralFoil / AeroSandbox (pinned)	fast section screening and trends (TR-06D control screen)	surrogate trends, no admission authority
SU2 (pinned)	sectional-CFD pilots (TR-05A1); later steady incompressible pilot on the disposition query	the only independent viscous lane — and it never admitted a coefficient (§6.3, §6.12)
DUST (pinned, Dockerized)	exact-geometry unsteady vortex-lattice/free-wake simulation: every screen, promotion, wake ladder, and diagnostic from TR-05D on	inviscid; no profile drag, stall truth, roughness, or acoustics
Repository runners (`scripts/aerodynamics/`)	deterministic batch generation, preregistration, validation, ledger, Studio serialization	the reproducibility spine

The toolchain was selected for **autonomy, not prestige**: the user required free, open-source tools an agent could install and run headlessly, with no paid services or GUI steps. Pinned runtimes and a DUST smoke path are committed under `sim/aero-toolchain/` with a third-party license inventory.

Equally important is the list of tools **considered but not used to produce any reported result**: BoTorch, pymoo, NSGA-II/III, CMA-ES, MAP-Elites, FLOWUnsteady, OpenVSP/VSPAERO, BladeAD, Xoptfoil2, and CST/Kulfan parameterization. They were evaluated and several remain good future options (§10), but the retained leader came from deterministic space-filling batches, complete local factorials, and frozen promotion rules. No genetic or Bayesian optimizer produced any candidate in this report, and describing the program otherwise would upgrade a brainstorm into an executed method.

4.4 Optimization strategy

The executed strategy, in the order it evolved:

1. **Deterministic space-filling first.** A 25-geometry DOE (§6.4), then a 24-geometry broad exact-DUST batch (§6.7), establish the landscape without optimizer path-dependence.
2. **Targeted acquisition over exhaustive atlases.** After uncertainty swamped ranking in the DOE, and after the sectional-CFD campaign's cost became clear, expensive computation was pointed at the single highest-value unresolved cell rather than a full grid (§5.3, §6.12).
3. **Causal microbatches.** Change one mechanism (incidence split, loop height) around a retained control to test a diagnosis, not to win (§6.6).
4. **Complete local factorials.** TR-05I through TR-06K are full 2×2×2 (or equivalent) crossings with preauthored candidate lists — no one-variable anecdotes, no post-hoc additions (§6.7–§6.11).
5. **Pareto-style non-dominated promotion** using thrust, power, cancellation, retention, manufacturability, and failure status — with at most **two frozen long-wake promotions per batch**.
6. **A preregistered stop rule:** two consecutive frozen batches with less than 2% mature-wake leader improvement and no wake/loading/structural/acoustic constraint fix end the local loop (§6.11). An analogous rule ended section-level refinement (§6.9).

4.5 Wake and convergence policy

Cold-start thrust is not a steady answer: an impulsively started free-wake solution carries startup loading that decays as the induced wake develops (§6.5 demonstrates this at scale). The policy that emerged:

- **Screens** run one or two revolutions and rank candidates *within a batch at the same window*. Screen numbers are never performance claims.
- **Promotions** run four uninterrupted revolutions under a frozen long-wake protocol.
- **The mature-wake extension gate** is evaluated per quantity (thrust and shaft torque) on the final revolution window, conjunctively: relative peak-to-peak ≤ 3%, |slope| per revolution ≤ 3%, and relative mean change from the prior revolution ≤ 2%. A locally flat final window can still fail — and for the retained design does fail — because the revolution-to-revolution mean is still moving or torque is not stationary (§6.12, Table 6).

The gate was never relaxed to let a result through, which is why every thrust figure in this report carries the phrase "wake gate closed."

4.6 Reproducibility and negative evidence

Five practices did the epistemic work:

1. **Freeze before looking.** Candidates, operating point, durations, solver identity, denominators, metrics, promotion rules, and stop conditions are committed before results are inspected.
2. **Explicit failure populations.** Diverged, non-finite, invalid-polar, out-of-domain, and nonstationary cases remain members of their populations: `valid + explicitFailures = eligible` is a checked invariant, and no batch drops a case to improve a rate (§6.2, §6.8's TR-05M and §6.11's TR-06K each retain two non-finite failures).
3. **Exact identities everywhere.** Geometry, mesh, protocol, source input, raw solver history, and cell-load artifacts are SHA-256-addressed.
4. **An immutable progress ledger.** `sim/rotor-progress/progress.jsonl` holds 39 append-only checkpoints (baseline, pending, rejected, improved, regressed, accepted) with `previousCheckpoint` links as the authoritative sequence; `PERFORMANCE.svg` is generated from it. The ledger exists because the user demanded a visible answer to "is the line going up or down" (§5.4) — it is a coordination instrument that became an evidence instrument.
5. **Method corrections stay visible.** The cross-platform mesh-byte mismatch, the restart-equivalence ambiguity, and the phase-misregistered flow probes are presented in §7 as corrections, not omitted implementation details.

This report adds one more layer: the paper-evidence extractor (`npm run aero:paper-data`) reads only the frozen sources, re-asserts the identity and denominator invariants on every run, emits the figure manifest and all ten generated figures deterministically, and byte-compares them in `--verify` mode. The figures in §6 cannot silently drift from the artifacts they cite.

---

5. Program narrative

The program ran across fourteen coordinated agent tasks — implementers, independent reviewers, a reporting/steering agent, a takeover agent, and an integration agent — with the user intervening at every turning point. The audited task index and full paginated transcript exports are in `docs/rotor-paper-kit/transcript-index.md` and `transcripts/`. Transcripts supply rationale and ordering here; every number is from §6's frozen sources. One audit caveat applies throughout: forked tasks inherit their parent's history, so the same statement appearing in several transcripts is duplication, not independent corroboration.

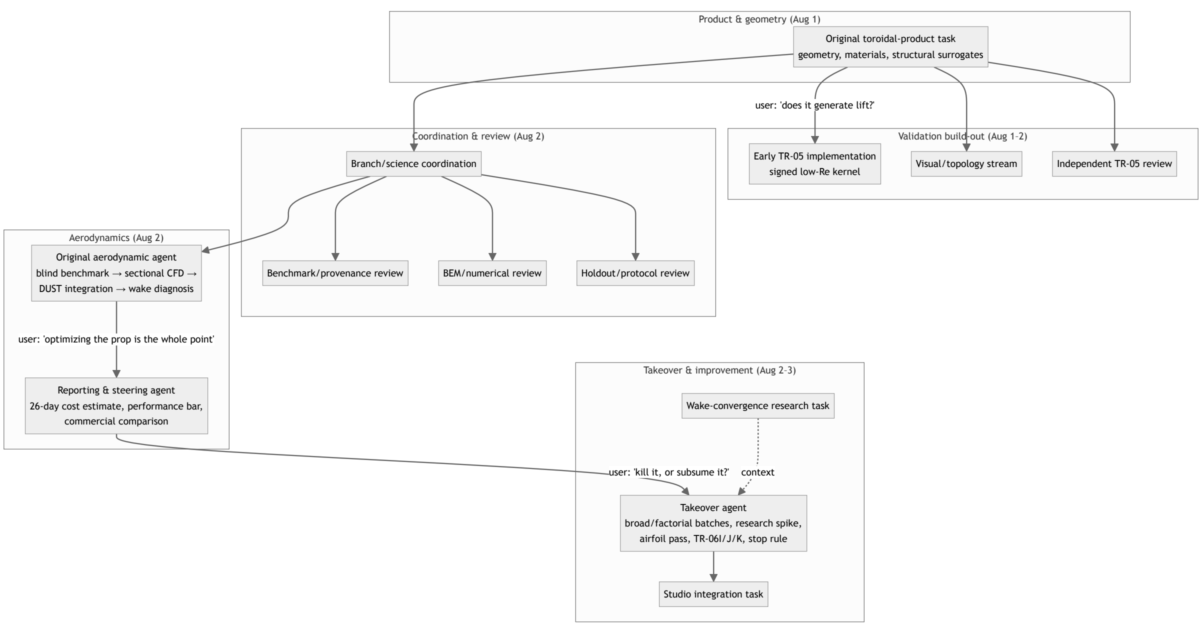


Figure 2 — task and handoff structure. Arrows show contextual and handoff relationships (with the user directive that drove each major transition), not sole authorship of commits.

5.1 A compelling object before credible lift (turning point 1)

The first task made the rotor a first-class product artifact: parametric geometry coupled to airframe clearance, analytic mass and inertia, structural surrogates, and material/process recommendations (commits `bee672d` through `3dbb9c4`). The physical stakes were understood early — ~114 m/s tip speed, ~30,000 g tip acceleration — and the material advice (PAHT-CF as a development candidate) was deliberately not a flight endorsement.

The geometry itself resisted. The first root cleanup reused a tip-oriented ellipse whose perimeter indexing mismatched the neighboring NACA rings: the loft joined a leading edge to a trailing edge, collapsed through an hourglass, and reopened with reversed winding. The tip was worse — a forced –90° keyframe and a keyframed point-loft folded the turnover through itself, with signed section area collapsing from ~2.50 mm² to 0.059 mm² and changing sign. The triangle generator faithfully rendered mathematically invalid surfaces, and higher resolution just made the wrong shape look more detailed. The durable correction was structural, not cosmetic: the turnover became a finite-radius fair sweep with transported section frames, continuous chord/twist laws, and explicit fairness/inversion gates (`13397c5`, `0b8dc42`). A separate incident — an apparent regression that was actually a stale `next start` build serving old code — taught the program that human visual review is only meaningful when the rendered artifact's commit and build identity are known.

The uncomfortable summary of this phase: a beautiful, increasingly well-engineered rotor existed, and none of it was lift evidence.

5.2 The user resets the order of operations (turning point 2 begins)

The pivotal user question was blunt: *is what you modeled actually supposed to generate lift yet or not?* The honest answer was "not defensibly." The critical path flipped from cosmetic refinement to signed loading, conventional controls, low-Re section coverage, uncertainty, and explicit failure states.

The early aerodynamic implementation made sound structural choices — signed force integration rather than clamped negative loads, explicitly non-lifting root blends, hard failure when a section query left its polar envelope, no fictitious 2-D lift on unresolved curved transitions, XFOIL polar provenance, and a conventional-propeller benchmark gate *before* any toroidal promotion. The first benchmark result was severe (§6.2): the model could not support performance claims, and no rotor-level correction could substitute for missing low-Re section physics. Commit `29194ce` and the first immutable checkpoint froze that failure as the starting line rather than hiding it. An independent review task confirmed the gap list: missing held-out evaluation, powertrain, uncertainty, and release gates.

5.3 Rigor becomes a local optimum (turning points 2–3)

The response to the benchmark failure was a preregistered SU2 sectional-CFD ladder — serious physics, frozen admission thresholds, and a planned grid of 3,600 qualification cells (§6.3). Its first full pilot diverged; its damped successor ran its entire 20,000-iteration budget without converging; CI exposed that mesh bytes differed between macOS and Linux. The responsible agent responded by hardening protocol identity, numerical ladders, and cross-platform contracts (`156e1b8`–`85986cf`) — careful, auditable, and increasingly disconnected from the objective.

The independent reporting task computed the operational consequence: at the measured pilot wall times, the serial campaign implied roughly two to four weeks of computation *before* medium meshes, transition cases, or URANS — and the first pilot had already concluded "URANS required, not admitted." The user supplied the governing correction: **optimizing the prop is the whole point.** Sectional CFD was demoted from prerequisite atlas to targeted information acquisition. This is a major methodological result in its own right: a workflow can obey every local rule of rigor and still optimize the wrong objective. Reproducibility and validation are constraints on useful experimentation, not substitutes for it.

5.4 Making progress visible

The same reporting task answered a second user demand — *show me whether the line is going up or down* — which began as a rebuke of caveat-walls and became infrastructure: a two-layer scorecard pairing a highly visible performance bar (hover, thrust-to-weight, commercial target) with the evidence class, denominator, and gate status needed to interpret it. Its descendant is the committed ledger and `PERFORMANCE.svg`.

The first 25-geometry DOE showed an internal signal (§6.4) — and the same scorecard immediately deflated it: the best resolved toroidal thrust was 18–28% of the measured HQProp control. The user's framing — the bar is *beating store-bought props*, not improving on a terrible toroidal baseline — became the fixed 2.8613 N target that every later figure carries.

5.5 The seductive wrong answer (turning points 4–5)

Commit `dc467dd` mapped the exact toroidal geometry into DUST with nanometer-scale mapping error and bit-for-bit deterministic reruns. The first load history reported ~4.089 N

— about 143% of the commercial bar. For a brief period the line looked spectacular.

The program did not declare victory. It preregistered a wake-duration ladder, and thrust decayed monotonically with simulated wake age: 3.728 N at a quarter revolution, 2.988 N at half, 2.008 N at one, 1.178 N at two (§6.5). Repeatability had proven only that the code repeated the same transient. The apparent breakthrough was startup loading before the induced wake developed, and the method rejected it (`b1e6194`). This episode is the program's strongest illustration of adversarial evidence: the system was designed to make exciting results work harder than ordinary ones, and it succeeded by invalidating the team's most exciting number.

Diagnostics then turned the decay into design information. After discarding a first, phase-misregistered probe set, phase-registered segment loads localized the loss: the rear system — rear leg, rear-leg blend, and rear-root blend — accounted for –0.494 N of the –0.719 N one-to-two-revolution thrust change, 68.7% of the loss (§6.6). A causal microbatch (incidence split and loop height only) improved the rear system without abandoning the smooth tip, supplying a mechanism, not a victory.

5.6 Subsumption (turning point 6)

By this point the original aerodynamic agent had again drifted toward fine-grained diagnostic work while the user expected candidate throughput. The user posed the collaboration question directly: *should we kill it, or subsume it?* The answer executed at commit `cd4dc23` was subsumption: freeze the original branch at a clean checkpoint, leave no solver running, preserve every protocol, hash, failure record, and runner — and continue with a new operator and an explicit throughput objective. The original agent's strict artifacts made the takeover possible; its execution strategy made the takeover necessary. Both halves of that sentence belong in the record.

5.7 Throughput changes the search (turning point 7)

The takeover ran 24 broad exact-DUST geometries with zero solver failures (`1255fab`). The one-revolution leader hit 2.554 N — 89.3% of the control — and immediately taught the next lesson: at two revolutions it fell to 1.59 N. A local factorial's two-revolution leader (1.7302 N) likewise fell to 1.4346 N at four. Short-wake ranking reverses (§6.7); promotion duration is part of the optimizer. Batches then became mechanism-directed — wake diversity (TR-05J), pitch×incidence crossover (TR-05K, `1142f1a`), load extension (TR-05L), and a solidity/retention factorial (TR-05M) — producing the pre-airfoil mature-wake baseline: 2.2952 N at revolution four, wake gate closed (§6.8).

5.8 The literature challenges the leader (turning point 8)

The research spike (`e0299e2`) did not reopen the topology question — the user had explicitly kept the smooth closed tips — but it audited the leader's physics and found the optimizer exploiting the model: zero of 48 lifting stations inside the attached-flow proxy domain, lifting thickness ratios of 27–54%, the curved tip transition carrying ~50% of reported load, and the rear leg nearly unloaded. It also performed the MIT correction (§2.1) and formalized the design-by-role model of §3.2. The high-pitch leader was retained *as evidence* but demoted as a physical design center. Proposed two-loop reseeds with wider sections improved plausibility on paper but were not carried through the final ladder — they remain hypotheses (§10), and this report does not claim them as results.

5.9 Bounded airfoil work (turning point 9)

At Re 20,000–40,000 the program froze a NACA-like control grid over camber, camber position, and thickness; screened it (TR-06D); correlated a bounded subset through pinned XFOIL (TR-06E, `045b263`); and kept manufacturability — physical minimum thickness, trailing-edge printability, leading-edge radius — inside the airfoil definition (TR-06G). SU2's independent viscous check remained blocked (`eb414c9`), so XFOIL trends were never promoted to rotor truth. A 6%-camber, 40%-position, 16%-thickness direction led; two local camber batches (TR-06H v1/v2) returned the same leader with zero robust-score improvement, firing the section stop rule (`e068fda`, `a71f4f8`). Only the *direction* — more camber within the tested neighborhood — transferred to the rotor.

5.10 Improvement, strikes, and stop (turning points 10–11)

TR-06I froze a complete 2×2×2 factorial (camber direction × 7.00/7.25 in pitch × −7/−8° incidence split), completed all eight screens, and promoted two candidates under the preauthored rule. The short-wake order reversed at four revolutions, and `factor-101-s3-pi7-risdm8` won: 2.4414 N, +6.37% over TR-05M, with +14.92% resolved shaft power and the wake gate still closed (§6.10, `cc69eba`). TR-06J (lower camber/pitch, more split) returned the unchanged anchor as leader — 0% improvement, strike one (`8d0b1e8`). TR-06K (bounded solidity reductions) completed six of eight screens, kept two non-finite outcomes as explicit failures, promoted only the anchor, reproduced 2.441372 N exactly in a third independent batch directory — 0%, strike two — and the preregistered rule fired (`97a15e2`, §6.11). Stopping was the planned outcome of a frozen rule, not a judgment call made after seeing results.

5.11 Honest integration and consolidation (turning point 12)

Commit `41bc4e9` integrated the exact retained rotor into Studio showing its 2.4414 N result, 85.3% control fraction, and **failed** wake status — the authoritative integration, superseding an earlier parallel-task integration that had exposed pre-TR-06 candidates. Commit `b508f68` consolidated the divergent rotor histories with a history-only merge, leaving the dirty monorepo `main` untouched. A supplementary research task on wake spin-up (6–8 + 2–3 revolutions as common practice for ordinary propellers) was recorded as context strengthening the four-revolution limitation, not as a retroactive rule.

After the handoff kit was frozen, three post-kit evidence events completed the record at this branch state: a fail-closed manufacturing-qualification surface (`149768e`, `1e7671f`); the TR-05 qualification disposition (`1f092aa`, `c881754`) that consolidated the retained candidate's evidence gap into one frozen report (§6.12); and a rejected steady-solver pilot on the disposition's selected query (`58cae16`, `5a7e337`), which kept the independent viscous lane explicitly closed. The first-print critical path (`rotor-first-print-critical-path.md`) froze this candidate as the sole article entering TR-09/#64 — an inert, inspected development print, with proof-spin, dynamometer, and flight explicitly out of its scope and no further local geometry optimization on the path.

5.12 Chronology summary

Phase	Date (2026)	Consequence
Product and geometry foundation	Aug 1	exact feasible geometry carrier; no lift evidence
Aerodynamic reality check	Aug 1–2	5/71 baseline coverage; claims impossible; failure frozen as baseline
Sectional-CFD detour	Aug 2	0 admitted cells; 12.6–26.1 serial days implied; redirected by user
Optimization pivot and toolchain	Aug 2	headless toolchain; DOE; commercial bar fixed at 2.8613 N
Exact DUST and wake reversal	Aug 2	4.089 N rejected; rear-system loss localized (68.7%)

Phase	Date (2026)	Consequence
Takeover and broad search	Aug 2	24/24 broad batch; short/long-wake reversals established
TR-05J–M	Aug 2	mature-wake baseline 2.2952 N, gate closed
Physics and airfoil reset	Aug 2–3	model-exploitation exposed; 6%-camber direction retained; section stop rule fired
TR-06I improvement	Aug 3	2.4414 N (+6.37% thrust, +14.92% power), gate closed
TR-06J/K stop	Aug 3	two 0% batches; two-strike rule fired
Integration and consolidation	Aug 3	Studio shows result with wake failure; histories consolidated
Post-kit qualification record	Aug 2–3 (local)	disposition frozen; steady pilot rejected; first-print path defined, print/flight locked

---

6. Results

Results are presented in the order the evidence matured, not in descending-thrust order. Every subsection names its frozen source; every population states its denominator; every thrust number carries its wake window and gate status. The whole-program checkpoint ledger is the connective tissue:

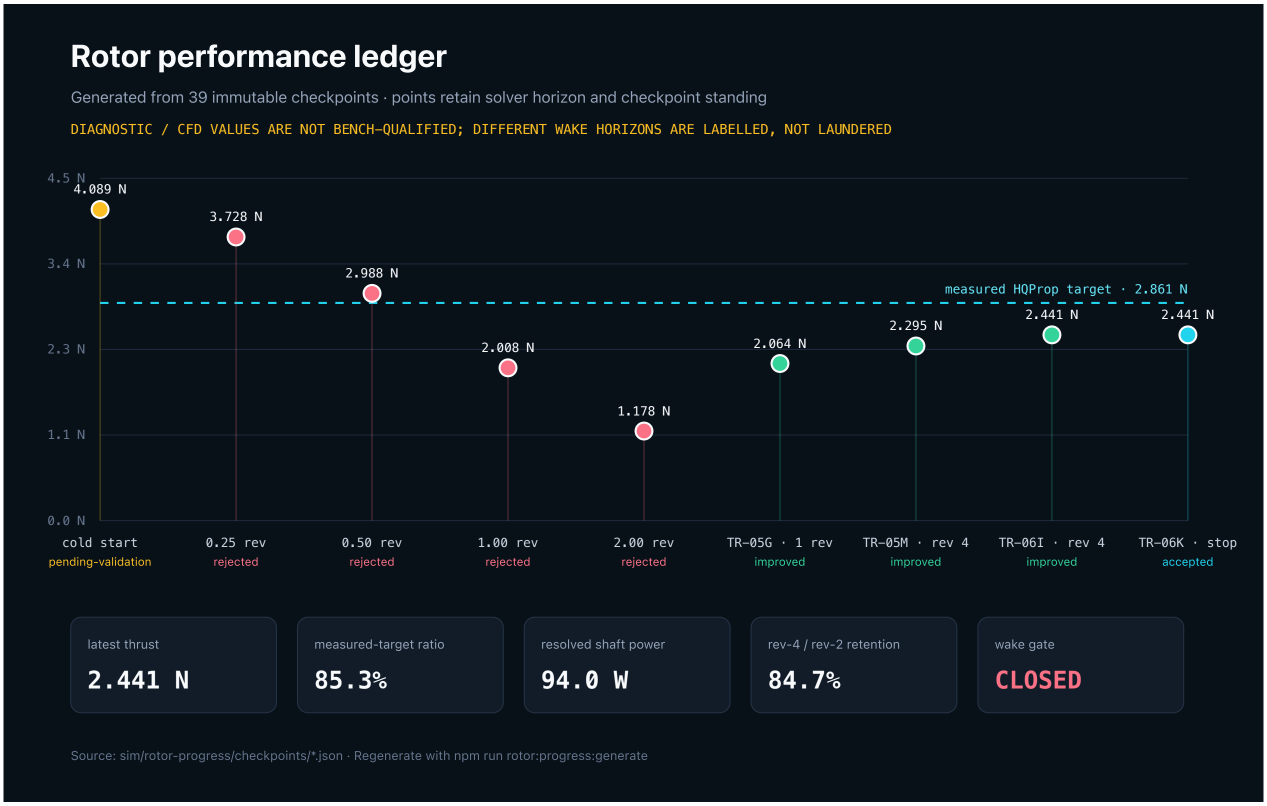


Figure 3 — the generated performance ledger (`sim/rotor-progress/PERFORMANCE.svg`), from the 39-checkpoint immutable ledger. Points differ in fidelity and wake duration; the figure encodes evidence level and status rather than implying one homogeneous time series. The dashed line is the 2.8613 N measured HQProp target.

6.1 Geometry feasibility

The exact three-loop smooth-tip family is feasible and deterministic: positive section area and consistent winding at every station, bounded frame rotation, finite sweep Jacobian, hub burial without bore or face escape, and audited minimum station thickness (1.284 mm for the retained design). Repeated generation is bit-identical, and identity hashing is stable across the engine and Studio serializations. This rung was earned through the failures of §5.1 — the invalid root and tip lofts are cataloged in §7 (F-01, F-02).

6.2 Conventional low-order model failure

The signed low-order model, judged against the frozen conventional held-out corpus with preregistered acceptance budgets (`tr-05a-v2/protocol.json`):

Population	Valid / eligible	Explicit failures
Former diagnostic corpus — corrected baseline	5 / 71	66
Former diagnostic corpus — bounded Snel/Liu candidate	6 / 71	65
Former diagnostic static rows — all candidates	0 / 13	13

On the rows that did resolve, error dwarfed every budget: CT median |relative error| 43.7% against a 10% budget, CT p95 75.1% against 15%, CQ median 32.3% against 15%, CQ p95 52.1% against 20%. **No conventional model was ever frozen**, and the one-shot holdout remains unopened and unspent.

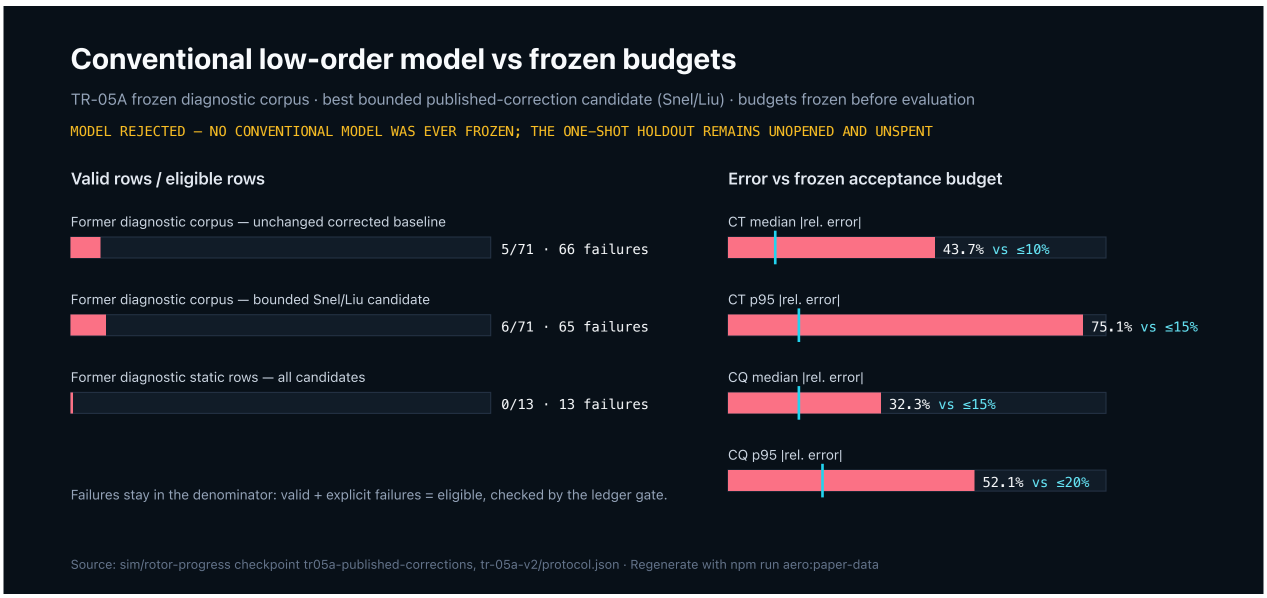


Figure 4 — coverage and error versus frozen budgets. Source: ledger checkpoint `tr05a-published-corrections` and the TR-05A-v2 protocol.

The earliest state resolved only 1 of 71 rows; the table above is the *best corrected* coverage after QPROP/circulation fixes and bounded published rotational corrections. The conclusion was structural: no rotor-level correction could substitute for missing low-Re transition, separation, and stall physics, which is what motivated both the sectional-CFD campaign (§6.3) and, ultimately, the move to exact-geometry DUST (§6.5).

6.3 Sectional-CFD cost and disposition

The preregistered SU2 sectional ladder (`condition-plan.json`) froze 3,600 qualification cells — 2 sections × 2 flow models × 9 Reynolds numbers × 4 Mach numbers × 25 angles of attack; operating-condition qualifications, not mesh elements. Executed: two pilots. Admitted coefficients: zero.

- The first full steady pilot **diverged** at inner iteration 8,859 (303.5 s wall time).
- An HLLC damped-ramp candidate ran its full 20,000-iteration budget (626.3 s) and failed every residual and force-stationarity gate — recorded as "URANS required, not admitted."

Multiplying the frozen cell count by the two measured wall times bounds the serial campaign at **12.6 to 26.1 days** — before medium meshes, transition models, or URANS. This derived bound, computed entirely from frozen data, is what the steering intervention of §5.3 acted on.

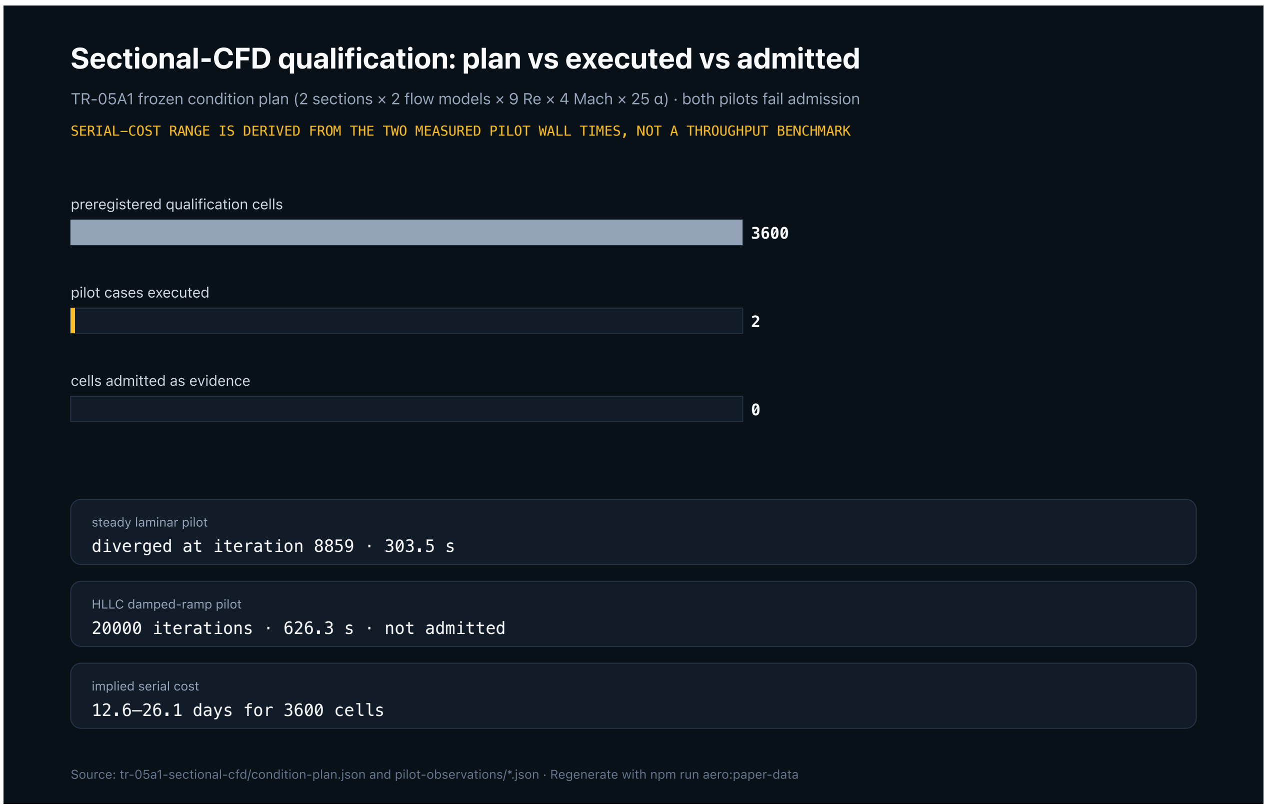


Figure 5 — the 3,600-cell plan, the two pilots actually run, and the derived serial-cost bound.

6.4 First DOE: internal signal, dominated by uncertainty

The first 25-geometry deterministic space-filling batch under the low-order model showed a real internal signal: `lhs-01` raised resolved hover-high thrust from the baseline's 0.114–0.177 N to roughly 0.521–0.811 N (~4.6× the resolved baseline), and `lhs-16` improved the conservative bound from −1.529 N to −0.615 N. But 23 of 24 valid candidates were non-dominated — uncertainty and unresolved turnover/interference loads overwhelmed ranking — and none was aerodynamically qualified. Against the measured control, even `lhs-01` reached only 18–28% of 2.8613 N. The "4.6× better than baseline" framing was retired in favor of the commercial bar (§5.4).

6.5 The cold-start result and the wake ladder

Exact-geometry DUST integration (nanometer mapping error, bit-identical reruns) produced a first-instant load of **4.089 N** — ~143% of the control bar, and bit-identical across two runs. The preregistered wake-duration ladder then measured trailing-mean thrust as the wake developed:

Wake duration	Trailing-mean thrust
first instant (t ≈ 0)	4.089 N
0.24 revolution	3.728 N
0.50 revolution	2.988 N
1.00 revolution	2.008 N
2.00 revolutions	1.178 N

Every longer window is lower; the two-revolution stage is explicitly flagged unstable; all five values were rejected as thrust evidence. The 4.089 N number appears in this report only as a rejected cold-start transient.

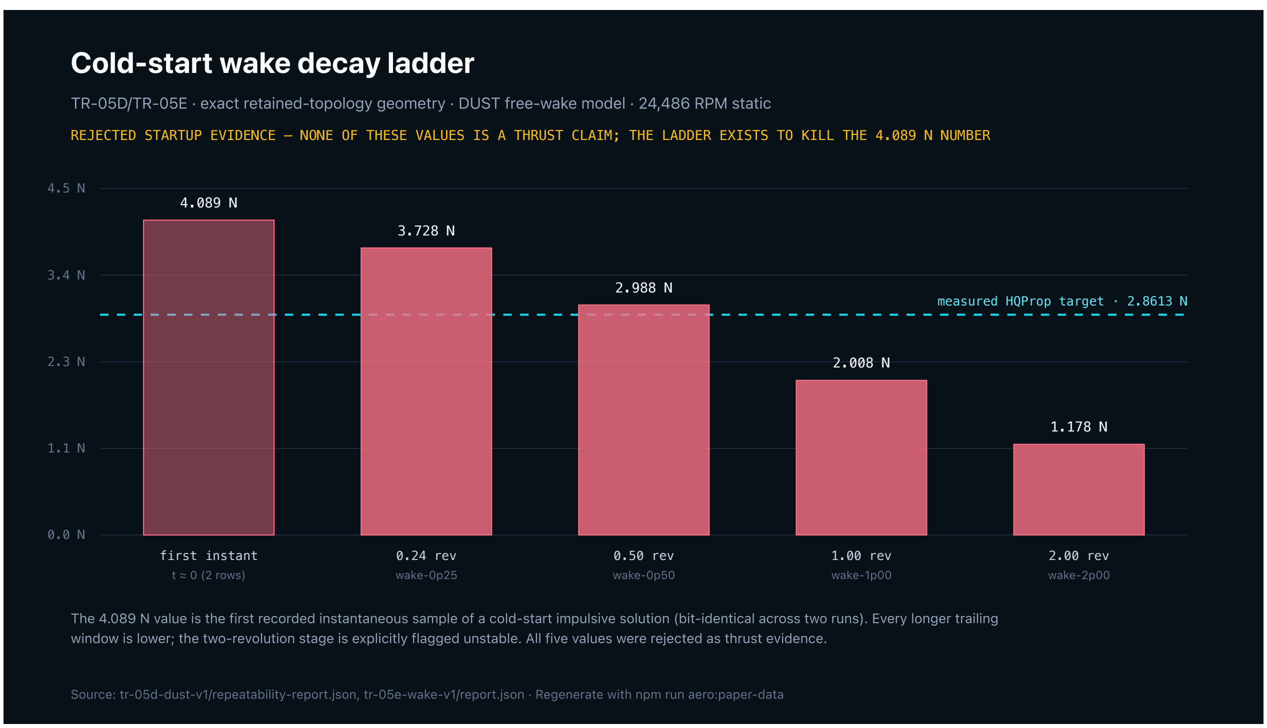


Figure 6 — the wake ladder that killed the apparent breakthrough. Sources: TR-05D repeatability report, TR-05E wake report.

6.6 Rear-system diagnosis

Phase-registered segment loads (TR-05F v2 — a first, phase-misregistered probe set was rejected; §7, F-09) decomposed the one-to-two-revolution thrust change of –0.7187 N. The rear system — rear-root blend (–0.242 N), rear leg (–0.157 N), and rear-leg-to-rear-root blend (–0.094 N) — contributed **–0.4936 N, 68.7% of the total loss**. (A live chat figure of "–0.399 N, ~56%" was a partial two-segment sum; the frozen report's three-segment definition is authoritative.) A causal microbatch varying only incidence split and loop height around the retained control (best: ≈ –4°, 6 mm, 2.064 N vs 2.008 N at one revolution) confirmed the mechanism was steerable without abandoning the closed tip.

Figure 7 — where the wake-development loss concentrates. All segments share one signed-load definition; red marks the rear system. Source: TR-05F v2 diagnosis block.

6.7 Broad search and the reversal lesson

The 24-geometry broad batch (TR-05H) completed 24/24 with zero failures. Its one-revolution leader `directed-02-high-pitch` reached 2.554 N — and fell to 1.594 N at two revolutions. The subsequent local factorial's (TR-05I) two-revolution leader 1.7302 N fell to 1.4346 N at four. These reversals are why every later batch separates screening windows from frozen four-revolution promotion, and why no screen number in this report is a performance claim.

6.8 TR-05J–M: the mechanism-directed ladder to the pre-airfoil baseline

Batch	Screen window	Population (completed / eligible, failures)	Screen leader	Screen thrust
TR-05H broad	1-rev trailing mean	24/24, 0	`directed-02-high-pitch`	2.554 N
TR-05I local factorial	1-rev trailing mean	8/8, 0	`local-p4p0-split-neg7-h7`	2.708 N
TR-05J wake diversity	2-rev mean	8/8, 0	`space-10`	1.684 N
TR-05K crossover ladder	2-rev mean	8/8, 0	`crossover-p5p4-split-neg7`	2.297 N
TR-05L load extension	2-rev mean	8/8, 0	`load-p7p0-front135`	2.660 N
TR-05M retention factorial	2-rev mean	6/8, **2 explicit failures**	`retain-rear80-root65-tip70`	2.694 N

Figure 8 — screening populations per frozen batch protocol. Non-finite solver outcomes stay in the denominator (TR-05M: 2, TR-06K: 2); no batch drops a case.

TR-05M's frozen four-revolution promotion produced the pre-airfoil mature-wake baseline: **2.2952 N** mean thrust, 81.827 W resolved shaft power, 85.19% fourth/second retention, 80.22% of the control target — wake gate closed.

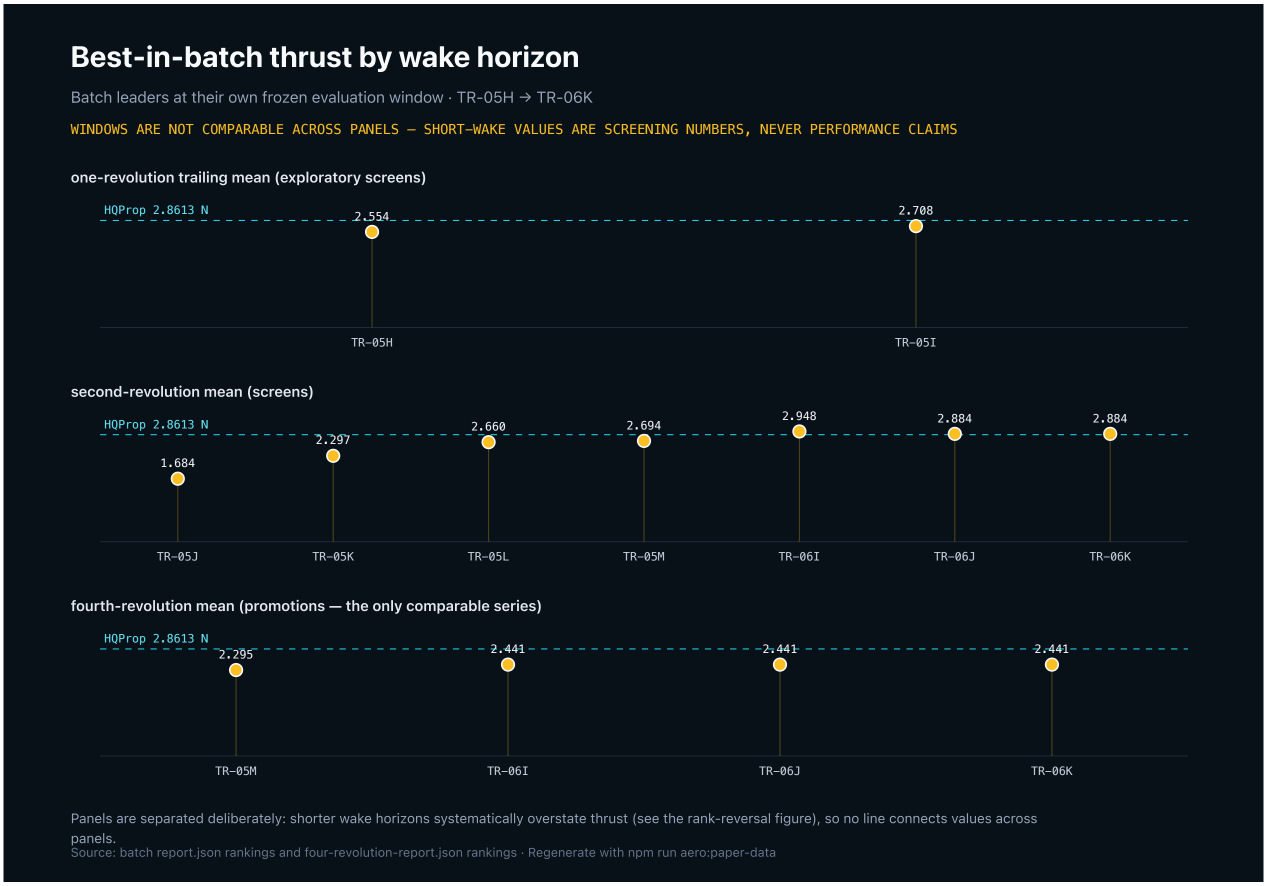


Figure 9 — batch leaders at their own frozen windows. Panels are deliberately disconnected: shorter horizons systematically overstate thrust, so no line connects values across windows, and only the four-revolution panel is a comparable series.

6.9 Airfoil screen and the blocked independent check

Within the bounded manufacturable XFOIL lane (frozen NACA-like grid over camber, camber position, thickness; pinned XFOIL; physical thickness and trailing-edge bounds inside the objective), a 6%-camber, 40%-position, 16%-thickness direction led. Two local camber batches returned the same full-coverage leader with zero robust-score improvement — the section stop rule fired. SU2's independent viscous confirmation remained blocked by numerical/mesh divergence, so **no XFOIL trend was promoted to rotor truth**; only the camber direction transferred. Committed per-batch charts: TR-06D, TR-06E, TR-06F (negative evidence), TR-06G, TR-06H v1/v2.

6.10 TR-06I: the retained improvement and its cost

TR-06I crossed the camber direction with pitch (7.00/7.25 in) and incidence split (–7/–8°) in a complete 2×2×2 factorial: 8/8 screens completed, two promotions under the preauthored rule. The two-revolution screen leader was `factor-011-s3-pi7p25-risdm7`

(2.948 N); at four revolutions the order reversed and `factor-101-s3-pi7-risdm8` won, 2.441 N to 2.428 N.

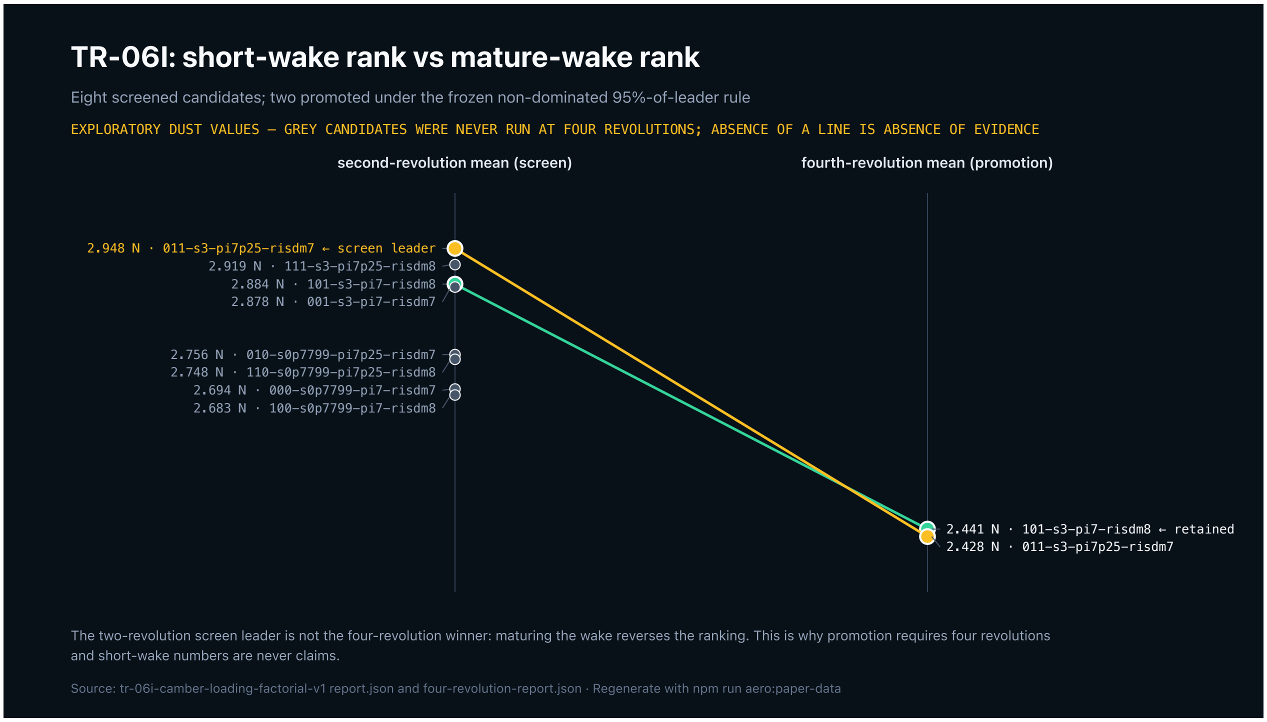


Figure 10 — the promotion-window reversal inside TR-06I. The screen leader is not the mature-wake winner; promotion requires four revolutions.

Against the TR-05M baseline at the same frozen four-revolution window:

Metric	TR-05M leader	Retained (TR-06I)	Change
Fourth-revolution mean thrust	2.2952 N	2.4414 N	**+6.37%**
Resolved shaft power	81.83 W	94.03 W	**+14.92%**
Thrust per shaft power	0.02805 N/W	0.02596 N/W	**–7.44%**
Rev-4 / rev-2 retention	85.19%	84.66%	–0.53 pt
Ratio to measured HQProp target	80.22%	85.32%	+5.10 pt
Full wake gate	closed	closed	not fixed

TR-05M → retained candidate: the whole tradeoff

Both leaders at the same frozen four-revolution DUST window, 24,486 RPM static

GREEN/ORANGE MARK DIRECTION OF CHANGE ONLY; THE WAKE GATE REMAINS CLOSED FOR BOTH DESIGNS

	TR-05M leader	retained (TR-06I)	change
fourth-revolution mean thrust	2.295 N	**2.441 N**	+6.37%
resolved shaft power	81.83 W	**94.03 W**	+14.92%
thrust per shaft power	0.02805 N/W	**0.02596 N/W**	−7.44%
rev-4 / rev-2 retention	85.19%	**84.66%**	−0.53 pt
ratio to measured HQProp target	80.22%	**85.32%**	+5.11 pt

what this is and is not

A reproducible raw mature-wake thrust gain bought with disproportionate shaft power — not an efficiency, commercial, acoustic, structural, or bench-qualified result.

Source: tr-05m and tr-06i four-revolution-report.json rankings · Regenerate with npm run aero:paper-data

Figure 11 — the whole tradeoff. A reproducible raw mature-wake thrust gain bought with disproportionate resolved shaft power — not an efficiency, commercial, acoustic, structural, or bench-qualified result.

6.11 TR-06J/K: diminishing returns and the stop

TR-06J (8/8 screens) and TR-06K (6/8 screens, 2 explicit non-finite failures) each promoted only the unchanged anchor, which reproduced **2.441372086606152 N and 94.032 W exactly** in the second and third independent batch directories. Improvement: 0% and 0%. Constraint fixes: none. The preregistered rule — two consecutive frozen batches under 2% mature-wake improvement with no constraint fix — fired, ending local rotor tuning.

Figure 12 — TR-06I's 6.37% gain followed by two 0% strikes. Stopping was the planned outcome of the rule, not a judgment call made after seeing results.

One design, three aliases: the anchor appears as `factor-101-s3-pi7-risdm8` (TR-06I), `factor-000-s3-pi7-risdm8` (TR-06J), and `factor-000-tcs0p65-bcs0p7-cs0p8` (TR-06K). The identical geometry and mesh hashes prove these are the same prop, not three winners; every figure and count in this report deduplicates by geometry identity.

6.12 The retained candidate and its evidence gap

The retained rotor — `factor-101-s3-pi7-risdm8`, geometry `558d516626af9df2391f2d898e7757d81d5844b91dde44ad31b064ebcbbc3e92`, mesh `827864e4102912583727575a06cfeb2b0b8e084082c3976e4c8a18a0c7d7a5b3` — is a 7-inch-pitch, –8° incidence-split, 6%-camber closed-tip design (full variable table in Appendix A). Its qualification standing, consolidated in the frozen TR-05 disposition report (`tr-05-qualification-disposition-v1/`):

- **Mature-wake gate: failed**, with a richer structure than a single number. Thrust passes relative peak-to-peak (1.59% ≤ 3%) and slope (0.31%/rev ≤ 3%) but fails mean-drift (4.93% > 2%); shaft torque fails all three criteria (3.23%, 3.19%/rev, 5.88%). The fourth-revolution window is locally flat and still not mature — exactly the failure mode §4.5's policy was designed to catch.
- **Operating envelope: 1 of 22 authored cells executed.** The envelope is 11 measured motor RPM coordinates × {static, 5 m/s axial}; only the exact 24,486

RPM static cell has run. The other 21 are explicit unresolved entries with null loads — the report never interpolates or scales the one result.

- **Aircraft coupling: central estimate only.** The frozen Studio build is 184.62 g AUW, needing 0.4526 N per rotor to hover. The DUST central estimate crosses hover and the 1.3/2.0 thrust-to-weight edges, but every conservative bound is null because wake, viscous, surface, current, and thermal uncertainty is unbounded.
- **Independent viscous evidence: rejected.** The disposition selected one bounded, decision-relevant sectional query — the extreme rear-leg cell (~51% thickness ratio, 35° incidence, Re ≈ 16k, outside the validated sectional regime). Its pinned incompressible laminar steady pilot ran all 4,000 iterations with a clean process exit and failed **every** residual (0/3) and force-stationarity (0/3) gate; no coefficient was admitted. The bounded next numerical experiment is a transition-aware URANS configuration — an experiment, not an authorization to tune the failed steady model.
- **Powertrain and mount: open.** Motor torque-speed equilibrium, current, and thermal coupling are not established; the 5 mm mount-hole spacing is an authored value with no measured counterpart.
- **Print and flight: locked.** `printOrFlightUnlockAllowed` is false; final disposition `unresolved-do-not-close-tr05`.

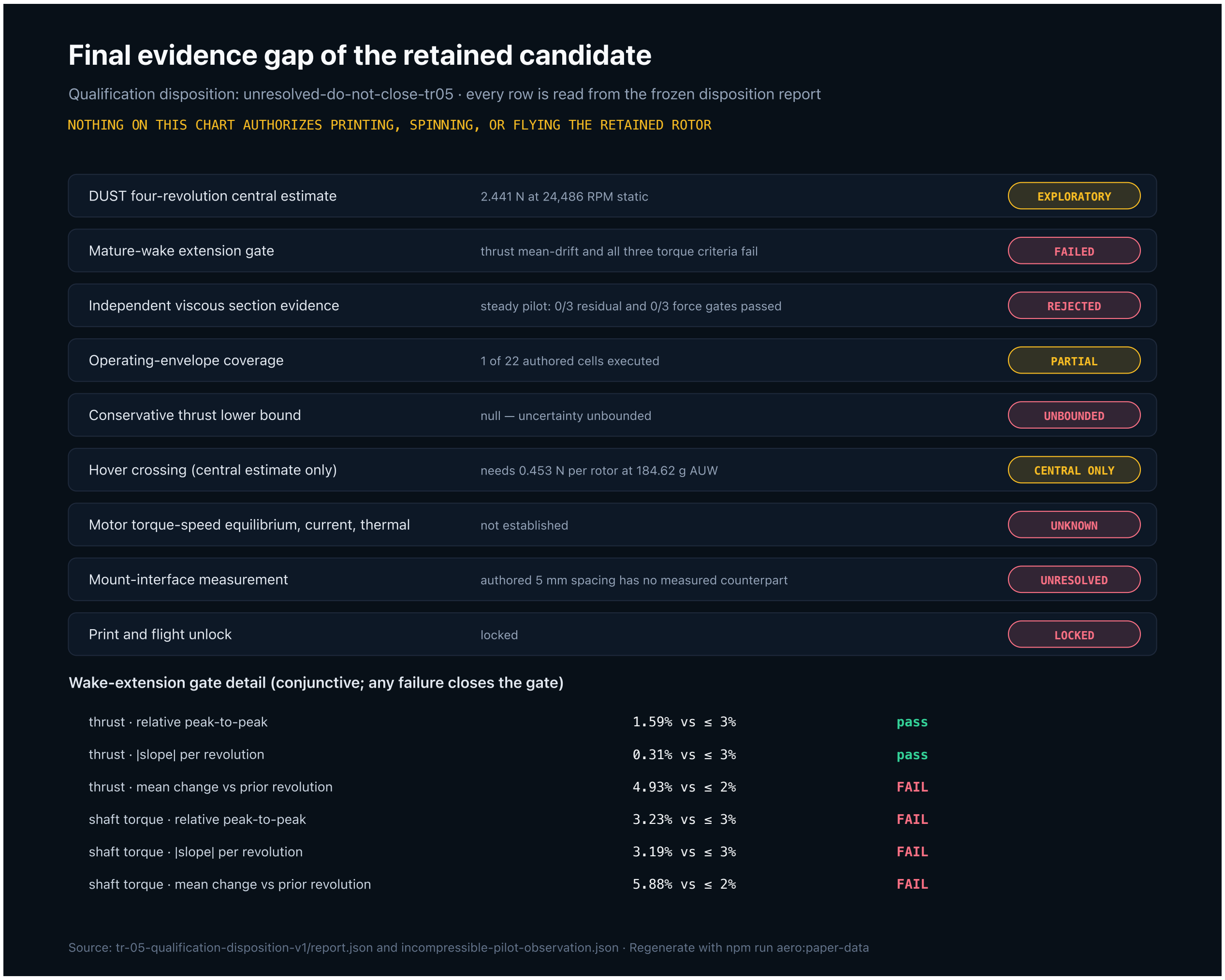


Figure 13 — the qualification matrix. Nothing on this chart authorizes printing, spinning, or flying the retained rotor.

The first-print critical path (`rotor-first-print-critical-path.md`) freezes this candidate as the sole article entering TR-09/#64 — one sliced, printed, conditioned, and *inspected inert development article*, explicitly not balanced, proof-spun, dynamometer-qualified, or flight-cleared — and places no further local geometry optimization on the path.

---

7. Failures and course corrections

This section has equal status with §6. Each entry records the observed symptom, the root cause, the corrective action, and — the field most often missing from engineering retrospectives — what evidence remained inadmissible afterward.

#	Observed symptom	Root cause	Corrective action	What sta[y] inadmiss
F-01	Root loft joined leading edge to trailing edge, collapsed through an hourglass, reversed winding	tip-oriented ellipse perimeter indexing mismatched neighboring NACA rings; tests encoded the same wrong assumption	correspondence-aware root blend; winding/area gates	all pre-fix ro geometry
F-02	Turnover folded through itself; signed section area fell 2.50 → 0.059 mm² and changed sign	forced –90° keyframe and keyframed point-loft; renderer faithfully drew invalid surfaces	finite-radius fair sweep with transported frames, continuous chord/twist laws, fairness/inversion gates (`13397c5` , `0b8dc42`)	all keyframe geometry ar renders
F-03	"Fixed" geometry appeared unfixed in review	stale `next start` production build serving old code	rebuilds, triangle counts, geometry contracts, exact URLs — build identity required for visual review	any visual approval wi a known bui identity
F-04	Visual/structural polish accumulating with no lift evidence	product momentum; plausible render mistaken for progress	user reset ("does it generate lift?"); critical path flipped to validation (§5.2)	every performance implication c pre-validatic artifact
F-05	Holdout data encrypted with no threat model	over-engineered blindness protocol	encryption removed after user objection (`19d1f0d`); one-shot output	none — the *protocol* sur minus the th

#	Observed symptom	Root cause	Corrective action	What sta[...] inadmiss[...]
			blindness retained	
F-06	1/71 initial held-out rows resolved; best corrected 6/71, 0/13 static; CT median error 43.7% vs 10% budget	missing low-Re transition/separation/stall section physics; no rotor-level correction can substitute	failure frozen as the baseline checkpoint (`29194ce`); sectional evidence made the next objective	every low-o[...] toroidal performanc[...] claim, permanentl[...]
F-07	Mesh bytes differed between macOS and Linux CI	non-deterministic cross-runtime serialization	versioned canonical mesh-byte contract with dual-platform proof (`bc358ef`); reviewer BLOCK held until proven	the numeric[...] checkpoint, the contract[...] existed
F-08	Sectional-CFD pilots: divergence at iteration 8,859; nonstationarity through 20,000 iterations; 3,600-cell plan implying 12.6–26.1 serial days	steady formulations inadequate for the regime; exhaustive-atlas strategy misaligned with the objective	campaign stopped by user steering; CFD demoted to targeted acquisition (§5.3)	all 3,600 pla[...] coefficients [...] zero were e[...] admitted
F-09	First turnover flow probes told an incoherent story	phase misregistration between compared runs	probes explicitly rejected; phase-registered v2 rerun (§6.6)	the v1 probe[...]
F-10	4.089 N — apparently	cold-start impulsive transient; repeatability ≠	preregistered wake ladder	the 4.089 N and all sub-[...]

#	Observed symptom	Root cause	Corrective action	What sta[ys] inadmiss[ible]
	143% of the commercial bar — repeatable bit-for-bit	validity	(§6.5); result rejected (`b1e6194`)	revolution th as evidence
F-11	Restart-equivalence ambiguity in wake continuation	unclear history-window semantics across restarts	uninterrupted four-revolution promotions; restart results quarantined	mixed restart/conti comparison
F-12	Short-wake leaders reversing at longer windows (2.554 → 1.594 N; 1.730 → 1.435 N; TR-06I screen leader losing at rev 4)	startup loading rewards high-cold-loading designs	frozen screen/promotion split; four-revolution promotion mandatory (§6.7, Figure 10)	every cross-window rank
F-13	Optimizer's favorite geometry exploiting the model: 0/48 stations in the attached-flow proxy domain, bridge carrying ~50% of load	inviscid panels reward curved-bridge loading no section evidence supports	literature/physics reset; design-by-role model; camber work moved to bounded manufacturable lane (§5.8)	the 7-inch b loaded lead a *physical* d center
F-14	SU2 independent section confirmation blocked; later	numerical/mesh divergence in the viscous lane at this regime	negative results retained; bounded URANS named as next	independen viscous validation, s absent

#	Observed symptom	Root cause	Corrective action	What sta inadmiss
	steady pilot 0/3 residual, 0/3 force gates		experiment — not run	
F-15	Two non-finite integral-load outcomes in TR-05M; two more in TR-06K	solver non-finite outcomes on aggressive candidates	retained as explicit failures in the denominator (Figure 8)	those four candidates
F-16	Zero improvement in TR-06J and TR-06K	local parameterization exhausted at current fidelity	preregistered two-strike stop rule fired (§6.11)	continuation local tuning productive a
F-17	No acoustic, structural, proof-spin, bench, or flight evidence of any kind	never attempted — correctly sequenced behind qualification	first-print critical path defined; print/flight locked	every quietr robustness, flightworthin claim

Three of these failures — F-06, F-08, and F-10 — each redirected the program. None was hidden, and every one left a frozen artifact. That is the practical meaning of "fail-closed": the failures are load-bearing parts of the record, not exceptions to it.

---

8. Discussion

8.1 The physical signal that is probably real

Two findings survive every caveat in this report as *mechanisms* (within the inviscid model). First, the rear system dominates wake-development loss — 68.7% of the one-to-two-revolution thrust change (§6.6) — which independently echoes published joined-tip and front/rear-interference results (§2.1). Second, that loss is steerable: incidence split, loop height, and the transferred camber direction each moved it, and the retained design's gain came with near-complete removal of negative rear-segment loading. Front/rear interaction management, not tip closure per se, is where this geometry's aerodynamic leverage lives.

8.2 The cost, stated plainly

The retained improvement is a raw thrust gain, not an efficiency gain: +6.37% thrust for +14.92% resolved shaft power is −7.44% thrust per shaft watt inside the same model. More pitch, camber, and loading bought thrust faster than the wake could pay for it. Any future claim of *efficiency* must come from a different design direction or a different fidelity — the current one moved the wrong ratio.

8.3 The bridge problem

The curved tip-bridge is the geometry's signature and its epistemic weak point. DUST's mean-camber panels will happily load it; no section evidence validates that loading; and §5.8 showed an optimizer left alone will migrate load onto exactly the surface the evidence covers least. The design-by-role discipline — treat the bridge as a structural/acoustic connector with modest net load until three-dimensional viscous evidence says otherwise — is the correct posture, and it is a *choice to forgo* apparent (possibly fictitious) performance.

8.4 Against the MIT mechanisms, carefully

MIT Lincoln Laboratory's materials hypothesize closed-tip stiffness, reduced tip-vortex contribution, and favorable thrust-at-a-given-sound in selected examples. This program neither reproduced nor refuted any of those claims: it has no acoustic data, no structural test, and an inviscid wake model. What it adds to the public conversation is narrower and still useful: at 88.9 mm and Re ≈ 20–40k, an exact closed-tip geometry can be brought within 85% of a measured commercial thrust bar *inside a free-wake model*, with the rear system identified as the binding aerodynamic constraint. Whether the hypothesized robustness and acoustic benefits justify the remaining thrust and power gap is precisely the question the physical test program (§10) exists to answer — and closed tips remain worth pursuing for those reasons even while the current thrust result trails the control.

8.5 The methodology is the transferable result

The program's most reusable artifact is the evidence-controlled optimization loop of §4.4–§4.6: freeze before looking; keep failures in denominators; give everything an exact identity; separate screening from promotion; pair every thrust with power and wake status; make stopping a preregistered rule; and treat agents as replaceable operators whose authority is always narrower than the ledger. Twice the program generated a result exciting enough to end less disciplined projects in celebration — 4.089 N, then the 2.948 N screen leader — and twice the machinery demoted it correctly. Once it generated a result boring enough to end less disciplined projects in abandonment — two 0% batches — and the machinery converted that into a reproducible, citable stopping decision. "Best tested toroidal hypothesis" and "best propeller" are different titles; the system never confused them.

8.6 On human-agent collaboration

Every turning point in §5 pairs an agent capability with a human judgment: agents supplied throughput, auditability, and tireless bookkeeping; the user supplied objective corrections that no amount of local rigor could generate ("does it generate lift?", "optimizing the prop is the whole point", "the bar is store-bought props", "kill it or subsume it?", "keep the smooth tips", "continue until diminishing returns"). The subsumption pattern (§5.6) is worth naming as a practice: when an agent's strategy stalls, freeze its verified assets at a clean checkpoint and change the operator and objective — punishing the agent by discarding its work would have cost the program its foundations.

---

9. Limitations and threats to validity

1. **Inviscid model.** DUST resolves no profile drag, separation, stall, transition, or surface roughness. All thrust, torque, and power figures are model quantities.
2. **The wake gate is closed.** Four revolutions left thrust mean-drift and all three torque criteria failing (§6.12); common practice for ordinary propellers is 6–8 spin-up plus 2–3 averaging revolutions, and this rotor's wake is not ordinary. The 2.4414 N figure is not a converged number.
3. **The control comparison is approximate.** The 2.8613 N bar is a manufacturer bench figure at matched diameter and RPM but different motor, voltage, and

measurement chain, with electrical-input (not shaft) power provenance. It is a necessary-condition bar, never a same-conditions experiment.

4. **No independent viscous evidence.** SU2 never admitted a coefficient in either the sectional campaign or the disposition pilot; XFOIL correlation is bounded and 2-D. The camber direction transferred on trend evidence alone.
5. **One executed envelope cell.** 21 of 22 authored operating cells are explicit unresolved nulls; nothing is known about off-design behavior.
6. **Unbounded uncertainty.** Every conservative aircraft-coupling bound is null; hover and thrust-to-weight crossings are central estimates only.
7. **No physical evidence.** No printed-surface, deformation, balance, FSI, thermal, dynamometer, acoustic, proof-spin, fatigue, or flight data exists. Robustness and quietness remain hypotheses.
8. **Screen-window fragility.** Short-wake rankings demonstrably reverse (§6.7, §6.10); any batch's screen ordering could differ at longer horizons, including the ones used to select promotions.
9. **Patent and freedom-to-operate questions** around toroidal propellers are outside this report's scope and unexamined.

---

10. Future work

Priority order follows a single principle from §6.11: change the evidence fidelity, not the local geometry.

1. **Stabilize an independent viscous lane.** Execute the bounded transition-aware URANS experiment on the disposition's selected rear-leg query; reproduce low-Re control sections before touching toroidal ones.
2. **Manufacture the exact frozen geometry** (TR-09/#64): record material, conditioning, slicer, orientation, mass, runout, and balance against the frozen identity; inspect before any rotation.
3. **Guarded proof-spin** behind containment with high-speed imaging and post-test inspection — before any thrust measurement.
4. **Matched-RPM dynamometer testing** with replicates and uncertainty: thrust, torque, electrical power, temperature, vibration, at matched diameter against the same commercial control.
5. **Calibrated acoustics at equal thrust:** narrowband spectra and directivity, separating tonal, broadband, and motor content — the first data that could ever

support the quietness hypothesis.

6. **Close the loop:** map measured pressure/deformation discrepancy back into the structural and aerodynamic models.
7. **Only then resume optimization** — multi-fidelity, with physical discrepancy and uncertainty inside the objective. The considered-but-unused stack (§4.3: BoTorch, pymoo, CST/Kulfan, two-loop reseeds) becomes relevant here, not before.

---

11. Conclusion

The program improved the credibility of the answer more than it improved the headline. Poltergeist now has a substantially better toroidal DUST candidate than it started with, exact geometry and mesh identities, frozen commercial controls, explicit failure records, an immutable 39-checkpoint ledger, and a bounded stopping rule that ended local tuning for stated, reproducible reasons. It does not yet have a toroidal rotor that beats the measured commercial control: the retained design increased mature-wake thrust 6.37% while losing 7.44% thrust per resolved shaft watt inside the same inviscid model, reached 85.32% of the measured bar, and left the full wake, viscous, acoustic, structural, and physical-test gates open. The rotor the program originally promised does not exist yet; what exists is prerequisite to it — an exact prop, a repeatable result, a method that can say no, and a map of the evidence that would change the answer. The next falsifiable test is already scheduled by the critical path: print the exact frozen article, inspect it, and — behind containment — spin it.

Appendix A — Retained candidate: identity, variables, reproduction

Field	Frozen value
Canonical ID	`factor-101-s3-pi7-risdm8`
Aliases (same	`factor-000-s3-pi7-risdm8` (TR-06J), `factor-000-tcs0p65-bcs0p7-cs0p8` (TR-06K)

Field	Frozen value
geometry hash)	
Geometry SHA-256	`558d516626af9df2391f2d898e7757d81d5844b91dde44ad31b064ebcbbc3e9`
DUST mesh SHA-256	`827864e4102912583727575a06cfeb2b0b8e084082c3976e4c8a18a0c7d7a5b`
Diameter / loops	88.9 mm (3.5 in) / 3
Global pitch	7 in
Front/rear incidence split	–8°
Maximum lifting camber	6% at the engine-native 40% position
Front chord scale	1.1240454412
Rear chord scale	0.8
Root-transition chord scale	0.65
Smooth tip-bridge chord scale	0.7

Field	Frozen value
Turnover-width scale	1.0763493634
Turnover-orientation bias	–4.3571050297°
Loop height	6 mm
Minimum audited station thickness	1.284 mm
Fourth-revolution DUST mean thrust	2.441372086606152 N
Fourth-revolution resolved shaft power	94.0319819788829 W
Rev-4 / rev-2 thrust retention	84.66%
Ratio to 2.8613 N measured HQProp target	85.32%

Field	Frozen value
Full mature-wake gate	**failed**
Evidence standing	exploratory, non-admissible diagnostic leader

The exact variables and identities are frozen in `tr-06i-camber-loading-factorial-v1/protocol.json` and repeated in the TR-06J/K protocols; the four-revolution results are in the respective `four-revolution-report.json` files, which reproduce the thrust and power figures bit-identically across three independent batch directories.

Wake-gate detail (final promotion window):

Quantity	Criterion	Value	Limit	Pass
Thrust	relative peak-to-peak	1.59%	≤ 3%	✅
Thrust	|slope| per revolution	0.31%	≤ 3%	✅
Thrust	mean change vs prior revolution	4.93%	≤ 2%	❌
Shaft torque	relative peak-to-peak	3.23%	≤ 3%	❌
Shaft torque	|slope| per revolution	3.19%	≤ 3%	❌
Shaft torque	mean change vs prior revolution	5.88%	≤ 2%	❌

Reproduction commands (from the repository root):

```
# Regenerate the paper-evidence manifest and all ten figures; fail on drift
npm run aero:paper-data
npm run aero:paper-data -- --verify

# Validate the immutable ledger and the paper kit's artifact manifests
npm run gate:rotor-progress
npm run aero:paper-kit

# Reproduce the qualification disposition and prepare the next bounded query
npm run aero:tr05-disposition
```

Per-batch reproduction commands are in each batch's `README.md` under `sim/toroidal-optimization/`; read `protocol.json` before `report.json`, and never infer long-wake standing from a short-screen report.

Appendix B — Data and figure provenance

Every figure numbered 4–13 above is generated by `scripts/build-rotor-paper-data.mjs` into `sim/toroidal-research/paper-v1/`, alongside `data.json`, the machine-readable manifest whose `sources` block lists all 24 frozen inputs with their SHA-256 hashes. The extractor asserts, on every run: the pinned 39-checkpoint ledger count; the retained candidate's geometry/mesh identities across TR-06I/J/K and the disposition; bit-equality of the four-revolution thrust across the three batch directories; consistency of the 2.8613 N control across the wake report, TR-05H, and the registry (including the g→N conversion); every population denominator (`valid + explicitFailures = eligible`); the 3,600-cell grid arithmetic; the improvement-ratio arithmetic against the ledger checkpoint; the closed wake gate; the rejected steady pilot; the locked print/flight state; and the rank-reversal premise (screen leader ≠ mature winner). Any drift fails the build rather than silently changing a figure. Figures 1–2 are Mermaid diagrams whose content is narrative structure, not data; Figure 3 is the independently generated ledger chart.

Evidence precedence, when sources could disagree: (1) immutable checkpoint JSON, (2) frozen protocol/report JSON, (3) geometry/mesh manifests and raw-output hashes, (4) generated ledger views, (5) git history, (6) methodology Markdown, (7) task-history exports — rationale only, never numbers.

Appendix C — Task record and transcript corrections

Fourteen agent tasks were audited page-by-page; the digest is `transcript-index.md` and the full exports are in `transcripts/` (user/assistant messages and turn metadata preserved; bulky outputs, diffs, tool arguments, and internal reasoning omitted). Forked tasks inherit parent history — repeated statements are not independent corroboration.

Task	Role
`019fbe04…` toridial prop	product concept, parametric geometry, materials, geometry repair, pivot to aerodynamic evidence

Task	Role
`019fbf0b-d64f…`	early signed low-Re kernel; preserved as partial work
`019fbf0b-f337…`	visual/topology stream; quarantined visual candidates
`019fbf0d…`	independent TR-05 review; documented the qualification gap list
`019fc24e…`	branch/worktree consolidation; scientific critical-path coordination
`019fc25e-ab29…`	benchmark/provenance review (provenance passed at `8e1b89f`)
`019fc25e-bda9…`	BEM/numerical review (BLOCK on cross-platform mesh identity)
`019fc25e-dab3…`	holdout/protocol review (fail-closed boundary verified)
`019fc2af…` prop aerodynamics	blind benchmark, sectional CFD, DUST integration, wake diagnosis; subsumed at `cd4dc23`
`019fc351…` reporting/steering	26-day cost estimate, performance bar, commercial comparison, toolchain vetting
`019fc382…` takeover	broad/factorial batches, research spike, airfoil pass, TR-06I/J/K, stop, integration, kit
`019fc439…` integration	Studio candidate surfaces; superseded by `41bc4e9`
`019fc528…` wake convergence	6–8 + 2–3 revolution context for the four-revolution limitation
`019fbe6d…` seance	unrelated migration monitor; branch-ownership context only

Live statements superseded by frozen evidence — preserved because the corrections are part of the method:

Live transcript statement	Frozen evidence	Treatment in this report
4.089 N is 143% of the commercial bar	wake ladder decays to 1.178 N at 2 rev	rejected cold-start transient (§6.5)
2.554 N broad leader at 89.3% of control	1.594 N at two revolutions	short-wake screen only (§6.7)
1.7302 N local leader improved the result	1.4346 N at four revolutions	rank-reversal evidence (§6.7)
rear system lost "–0.399 N, ~56%"	–0.4936 N, 68.7% (three-segment frozen definition)	frozen figure used; chat figure noted as partial sum (§6.6)
TR-05M 2.295 N called "mature-wake leader"	wake gate closed	longer-window diagnostic baseline (§6.8)
TR-06I "accepted" in checkpoint decision	acoustic/structural/viscous/bench/wake absent	accepted *stopping decision*, not accepted product (§6.11)
genetic/Bayesian methods recommended	leader from deterministic factorials	considered-but-unused (§4.3)

Appendix D — Claims matrix

Claim	Status	What would change it
The exact closed-tip geometry is feasible, deterministic, printable-in-principle	**Supported** (geometry gates, identities)	—
The retained rotor produces 2.4414 N at 24,486 RPM static in the frozen four-revolution DUST lane	**Supported as a model result**, wake gate failed	longer-horizon runs; independent viscous model
The retained rotor improved mature-wake thrust 6.37% over TR-05M	**Supported within the model**, at +14.92% power	—
The retained rotor beats the commercial control	**Unsupported** — 85.32% of the bar	matched-RPM dynamometer measurement
The retained rotor is more efficient	**Unsupported and contradicted** — −7.44% N/W in-model	a design/fidelity change that moves the ratio
The wake is mature / the gate passed	**False** — Appendix A table	more revolutions and a passing gate
The rotor is quieter	**Hypothesis** — no acoustic data	calibrated spectra at equal thrust
The rotor is structurally robust / flightworthy	**Hypothesis** — no physical article exists	print, inspection, proof-spin, endurance
Independent viscous validation exists	**False** — 0 admitted coefficients, steady pilot rejected	a converged, gate-passing URANS/section result
MIT published peer-reviewed proof of a universal toroidal advantage	**False** — patent + technology summary only	—
Genetic/Bayesian optimization produced the candidate	**False** — deterministic factorials did	—

Appendix E — Key commits

Commit	Role
`bee672d`	preserve partial low-Re screening work without performance claims
`3dbb9c4`	approved rounded root geometry; aerodynamic evidence becomes critical path
`13397c5`, `0b8dc42`	replace broken tip loft with finite-radius fair turnover
`292e9ae`	preregister blind aerodynamic validation
`19d1f0d`	remove unnecessary holdout encryption
`29194ce`	freeze corrected-but-inadequate conventional baseline
`2f62eae`	add the immutable rotor-progress ledger
`156e1b8`	preregister the sectional-CFD harness
`8e1b89f`	record the 20,000-iteration nonstationary candidate
`bc358ef`	stabilize mesh identities after cross-runtime mismatch
`3ea111b`	seed the first deterministic toroidal optimization loop
`1515e12`	add the headless aerodynamic toolchain
`dc467dd`	integrate exact geometry with DUST
`b1e6194`	reject the first wake ladder rather than accept the cold-start number
`4749b0f`	diagnose rear-system wake loss with phase-registered probes
`66848dc`, `cd4dc23`	causal microbatch; original-agent handoff checkpoint
`1255fab`	24-geometry takeover broad search
`1142f1a`	crossover mechanism result
`37a130f`, `7d79b79`	establish and checkpoint the TR-05M retention leader
`a6a22e1`	record the reusable optimization methodology

Commit	Role
`e0299e2`	physics/literature reset for low-Re toroidal design
`045b263`	pinned XFOIL section correlation
`eb414c9`	retain the blocked independent SU2 result
`e068fda`, `a71f4f8`	section diminishing-return stop
`cc69eba`, `6c235e4`	record and freeze the TR-06I mature-wake improvement
`8d0b1e8`, `8e90aea`, `97a15e2`	strikes, final reproduction, rotor stop
`41bc4e9`	integrate the exact retained leader into Studio
`b508f68`	consolidate rotor research histories
`68ddbaf`, `0146888`, `04e4151`	paper plan, handoff kit, transcript exports
`c80f382`	reconcile the first-print critical path (TR-09/#64)
`149768e`, `1e7671f`	fail-closed manufacturing qualification (engine + Studio)
`1f092aa`, `c881754`	TR-05 qualification disposition and final checkpoint
`58cae16`, `5a7e337`	record and checkpoint the rejected steady pilot