Predictive Geometry-Only Rotation Curves: RFT v2 vs Fair k=0 Baselines

RFT Cosmology Project[1]

[1]https://rft-cosmology.com

November 11, 2025

Abstract

Predictive models of galaxy rotation curves without per-galaxy dark matter tuning remain a key test for alternative gravity frameworks. We present RFT v2, a geometry-only resonance solver with six global parameters and zero per-galaxy tuning ($k = 0$), evaluated on a blind TEST cohort of 34 SPARC-derived disk galaxies.

RFT v2 achieves **58.8% pass@20%** (20/34) compared to 52.9% for the fair global NFW baseline and 23.5% for canonical MOND. Paired McNemar tests show that RFT is **competitive with NFW** ($p = 0.69$, overlapping Wilson intervals) and **significantly ahead of MOND** ($p = 0.004$). The predictive delta versus NFW is 5.9 percentage points, reinforcing the need for larger cohorts to establish significance.

Beyond aggregate metrics, the fairness pack and robustness suite highlight the intended mechanisms. RFT is the only k=0 model with non-zero low-surface-brightness wins (Table 4), consistent with the acceleration-gated tail activating where baryonic gravity is weak. $\pm 10\%$ parameter sweeps leave the pass rate unchanged (Figure 4), and ablations plus the rotation-curve gallery (Figures 5 and 6) confirm that the tail term is the dominant causal lever while unveiling representative wins, ties, and failures.

We release a complete reproducibility pack: `RUNME.sh`, CI gates that enforce both the baseline lock and the new numbers hash lock, frozen configs at tag `rc-v2-green-20pct` (commit `3428db0f`), and the figures/tables generated in this paper. The result is a transparent, parameter-efficient benchmark that isolates what has been demonstrated on SPARC-99 and what still remains open.

1 Introduction

Galaxy rotation curves remain a stress test for any theory that seeks to explain the flattened outer profiles of spiral galaxies without an ad hoc halo tuned to each system. While ΛCDM fits typically invoke two halo parameters per galaxy and MOND introduces a low-acceleration law, direct predictive comparisons across paradigms rarely enforce matched parameter budgets. As a result, literature claims often contrast a predictive model with a descriptive one, obscuring whether a purported advantage comes from new physics or from extra degrees of freedom.

We frame the problem explicitly in terms of the per-galaxy parameter budget k. Predictive settings fix $k = 0$—all parameters are learned on TRAIN and frozen on blind TEST—whereas descriptive settings allow $k \geq 2$ per galaxy. Our baselines therefore include three k=0 models (RFT v2, global NFW, MOND) plus a k=2 reference (per-galaxy NFW fits) that represents the descriptive ceiling but is not a fair competitor. Table 3 summarises these budgets and the degrees of freedom allocated to each solver.

RFT v2 extends the geometry-only kernel developed in earlier RFT campaigns with a frozen acceleration-gated tail and an identity convolution kernel, yielding six global parameters and zero per-galaxy tuning. On the blind TEST cohort (34galaxies) this predictive configuration passes the 20% RMS criterion in 20/34galaxies (58.8%), compared with 18/34(52.9%) for the global NFW halo and 8/34(23.5%) for canonical MOND. The paired McNemar test indicates that the RFT vs NFW difference is suggestive but *not* statistically significant ($p = 0.69$), whereas the RFT vs MOND gap is highly significant ($p = 0.004$). These results therefore motivate honest framing: RFT is competitive with a fair NFW baseline and significantly ahead of MOND, but it does not yet deliver a decisive win over ΛCDM.

This paper is scoped strictly to the galaxy rotation campaign. The L-Series cluster lensing program reported a negative predictive result in its final audit, so no claims about lensing or cosmological-scale performance are made here; those tracks remain separate until they achieve the same level of rigor. Our goal is to document what is and is not established by the SPARC-99 study and to freeze the accompanying data/figure assets for peer review.

Our contributions are: (i) a transparent predictive comparison between RFT v2 and k=0 baselines with matched parameter budgets; (ii) a fairness pack that enumerates head-to-head wins, Wilson confidence intervals, and LSB/HSB behavior; (iii) robustness evidence via ±10% parameter perturbations, ablations, and representative rotation-curve galleries; and (iv) complete reproducibility assets (RUNME, CI gates, and hash-locked numbers) that allow anyone to regenerate the figures and tables in Sections 3–4. The remainder of the paper proceeds as follows: Section 2 details the dataset, metrics, and computation pipeline; Section 3 presents the fairness, robustness, and gallery analyses; Section 4 interprets the findings and limitations; and Section 5 provides conclusions and future work.

2 Methods

2.1 Data and Cohort Selection

We use a SPARC-derived dataset [?] of 99 disk galaxies with high-quality rotation curve measurements. The cohort was split into TRAIN ($n = 65$) and TEST ($n = 34$) subsets *a priori*, with TEST galaxies held blind during model calibration. All results reported in this work use the TEST cohort exclusively.

Galaxy selection criteria:

- High-quality photometry (Spitzer 3.6 μm)
- Well-sampled rotation curves (minimum 10 independent radial bins)
- Reliable distance estimates
- Velocity uncertainties $< 20\%$

TRAIN and TEST manifests are provided as frozen artifacts (`cases/SP99-TRAIN.manifest.txt`, `cases/SP99-TEST.manifest.txt`) to ensure reproducibility.

2.2 Performance Metric

A galaxy is considered to **pass@$X\%$** if the root-mean-square (RMS) percentage error between model and observed rotation velocities is $\leq X\%$:

$$\mathrm{RMS\%} = 100\% \times \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{v_{\mathrm{model},i} - v_{\mathrm{obs},i}}{v_{\mathrm{obs},i}}\right)^2} \tag{1}$$

where N is the number of radial bins, $v_{\mathrm{model},i}$ is the predicted circular velocity at radius r_i, and $v_{\mathrm{obs},i}$ is the observed velocity with uncertainty σ_i.

We report two thresholds:

- **Pass@20%** (primary): Tolerates modest ($\leq 20\%$) systematic deviations
- **Pass@10%** (secondary): Stricter criterion for high-precision fits

The 20% threshold is defensible for heterogeneous disk galaxies with varied morphologies, inclinations, and distance uncertainties [**?**]. We also report median RMS% and Wilson 95% confidence intervals on pass rates.

2.3 RFT v2 Model

2.3.1 Tail Formula

The RFT v2 geometry-only solver applies an acceleration-gated tail term to the Newtonian baryonic rotation curve $v_{\mathrm{bar}}(r)$:

$$v^2_{\mathrm{RFT}}(r) = v^2_{\mathrm{bar}}(r) + r \cdot g_{\mathrm{tail}}(r) \tag{2}$$

where the tail acceleration is:

$$g_{\mathrm{tail}}(r) = A_0 \left(\frac{r_{\mathrm{geo}}}{r}\right)^{\alpha} \cdot \frac{1}{1 + \left(\frac{g_b}{g_*}\right)^{\gamma}} \cdot \left[1 - \exp\left(-\left(\frac{r}{r_{\mathrm{turn}}}\right)^{p}\right)\right] \tag{3}$$

Components:

- A_0 [km^2 s^{-2} kpc^{-1}]: Amplitude
- α [-]: Power-law slope (controls how tail scales with radius)
- r_{geo} [kpc]: Geometric scale (from RFT resonance framework; here taken as galaxy-specific from baryonic mass distribution)
- g_b [km^2 s^{-2} kpc^{-1}]: Baryonic acceleration at radius r
- g_* [km^2 s^{-2} kpc^{-1}]: Acceleration threshold (gate activates when $g_b \ll g_*$)
- γ [-]: Gate steepness
- r_{turn} [kpc]: Radial onset scale (suppresses inner disk)
- p [-]: Onset exponent

Design intent: The acceleration gate $[1 + (g_b/g_*)^{\gamma}]^{-1}$ activates the tail boost where baryonic gravity is weak ($g_b \ll g_*$), naturally favoring low surface brightness (LSB) galaxies. The radial onset $1 - \exp(-(r/r_{\mathrm{turn}})^p)$ suppresses the tail in the inner disk, allowing outer rotation curve features to dominate.

Identity kernel: We use an identity convolution kernel (no smoothing) to isolate the tail physics without confounding from spatial averaging.

2.3.2 Frozen Parameters

All six global parameters were calibrated on the TRAIN cohort ($n = 65$) using a bounded grid search with BIC-based selection. The frozen configuration (Table ??) was then evaluated once on the blind TEST cohort ($n = 34$) with zero per-galaxy tuning ($k = 0$):

Table 1: Frozen RFT v2 Parameters (Tag: `rc-v2-green-20pct`, Commit: `3428db0f`)

Parameter	Symbol	Value
Amplitude	A_0	1000 km^2 s^{-2} kpc^{-1}
Tail slope	α	0.6
Accel threshold	g_*	1000 km^2 s^{-2} kpc^{-1}
Gate steepness	γ	0.5
Radial scale	r_{turn}	2.0 kpc
Onset exponent	p	2.0

2.4 Baseline Models

2.4.1 NFW_global (k=0)

A Navarro-Frenk-White (NFW) dark matter halo [?] with *two global parameters* shared across all TEST galaxies:

$$\rho_{\text{NFW}}(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1 + \frac{r}{r_s}\right)^2} \tag{4}$$

where ρ_s (density scale) and r_s (scale radius) are fixed globally. This provides a fair $k = 0$ comparison (no per-galaxy tuning).

2.4.2 MOND (k=0)

Modified Newtonian Dynamics [1] with the canonical acceleration scale $a_0 = 1.2 \times 10^{-10}$ m s^{-2} (one global parameter):

$$v^4_{\text{MOND}} = v^4_{\text{bar}} a_0 r \tag{5}$$

This is the simplest MOND formulation; more sophisticated interpolation functions exist but add free parameters.

2.4.3 NFW_fitted (k=2, Reference Only)

For context, we also report per-galaxy NFW fits with *two parameters per galaxy* (ρ_s, r_s fitted individually). These descriptive fits typically clear ~80% pass@20% on TEST but require 68 total parameters (34 galaxies × 2), making them a **descriptive** rather than predictive baseline. We present them only to clarify the $k = 0$ vs $k = 2$ paradigm distinction.

2.5 Training Protocol

2.5.1 Grid Search and Selection

RFT v2 parameters were selected via bounded grid search on TRAIN ($n = 65$):

1. **Grid bounds**: Physically motivated ranges for each parameter (e.g., $A_0 \in [500, 2000]$, $\alpha \in [0.3, 1.0]$)

2. **Selection rule**: Minimize Bayesian Information Criterion (BIC) on TRAIN:

$$\text{BIC} = N \ln(\text{RSS}/N) + k \ln(N) \tag{6}$$

   where RSS is residual sum of squares, N is total data points, $k = 6$ (global parameters)

3. **Stop rule**: Pre-registered; no post-hoc tuning after freeze

4. **Freeze**: Best configuration locked via git tag (`rc-v2-green-20pct`, commit `3428db0f`)

2.5.2 Pre-Registration

Grid bounds, selection criteria, and TEST evaluation protocol were documented in `RFT_V2.1_PREREG.md` before TEST evaluation. This ensures honest reporting and prevents p-hacking.

2.6 Statistical Tests

2.6.1 Primary Test: McNemar's Exact (Paired)

Because the *same 34 TEST galaxies* are evaluated by all models, we use **McNemar's exact test** to assess pairwise differences in pass/fail outcomes. For two models A and B:

- Let b = number of galaxies where A passes and B fails
- Let c = number of galaxies where B passes and A fails

Under the null hypothesis (H_0: marginal probabilities equal), b follows a Binomial($b + c$, 0.5) distribution. The two-sided p-value is:

$$p = \sum_{k: P(X=k) \leq P(X=b)} P(X = k), \quad X \sim \text{Binomial}(b+c, 0.5) \tag{7}$$

This is the **correct test for paired binary outcomes** and is reported as the primary statistical result in this work.

2.6.2 Secondary Test: Two-Proportion z-Test (Unpaired)

For comparison, we also report the unpaired two-proportion z-test in the Methods section (not as a primary claim):

$$z = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}, \quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \tag{8}$$

where p_1, p_2 are observed pass rates and x_1, x_2 are pass counts. This test *ignores pairing* and is less conservative; we include it for methodological transparency but do not base claims on it.

2.6.3 Wilson Confidence Intervals

For binomial proportions (pass rates), we use Wilson score intervals [?] rather than normal approximation, as they are more accurate for small n and near-boundary proportions:

$$\text{CI} = \frac{1}{1 + \frac{z^2}{n}} \left(\hat{p} + \frac{z^2}{2n} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}} \right) \tag{9}$$

where $z = 1.96$ for 95% confidence, $\hat{p}$ is the observed pass rate, and n is the cohort size.

2.7 Reproducibility Protocol

All analysis is fully reproducible via a one-click verification script:

```
./RUNME.sh
```

This executes:

1. Baseline consistency audit (Gate 0): Verifies frozen numbers match publication
2. Final numbers hash check (Gate P1): Ensures statistical results unchanged
3. Fairness pack generation: Computes McNemar tests, Wilson CIs, LSB/HSB splits

4. Stability analysis: Runs ±10% perturbations
5. Figure regeneration: Produces all camera-ready plots from frozen JSON

CI/CD workflows (GitHub Actions) enforce these gates on every commit, preventing accidental number drift. All source code, data manifests, and frozen configurations are available under MIT license at:

https://github.com/rft-cosmology/rft-v2-galaxy-rotations

SHA256 checksums of all frozen artifacts are provided in `rft-v2-repro-1.0.SHA256SUMS.txt`.

2.8 Code-to-Paper Mapping

For full traceability, we provide a mapping between mathematical expressions and implementation:

Table 2: Code-to-Math Mapping (Reproducibility)

Paper Symbol	Code Variable	File
A_0	A0_kms2_per_kpc	config/global_rc_v2_frozen.json
α	alpha	config/global_rc_v2_frozen.json
g_*	g_star_kms2_per_kpc	config/global_rc_v2_frozen.json
γ	gamma	config/global_rc_v2_frozen.json
r_{turn}	r_turn_kpc	config/global_rc_v2_frozen.json
p	p	config/global_rc_v2_frozen.json
Pass@20%	pass_20_rate	results/.../test_results.json
RMS%	rms_percent	results/.../test_results.json

Exact formulas, grid bounds, and selection rules are documented in:

- Model implementation: `solver/rft_v2_gated_tail.py`
- Grid search: `scripts/grid_search_v2.1_refine.py`
- Metrics: `metrics/compute_metrics.py`
- Statistical tests: `analysis/fairness/compute_paired_stats.py`
- Fairness pack (head-to-head + LSB/HSB): `scripts/generate_fairness_pack.py`
- Stability perturbations: `scripts/generate_stability_analysis.py`
- LaTeX macro emission: `scripts/emit_tex_numbers.py`

2.9 Reproducibility and Provenance

RUNME pipeline: `RUNME.sh` verifies the solver build, regenerates fairness/stability JSON artifacts, emits `paper/build/numbers.tex`, and run-time checks the CI gates. The GitHub Actions workflow enforces two guardrails on every commit: (i) the baseline lock via `scripts/audit_baselines.py` and `scripts/verify_final_numbers_hash.py` (expected SHA256 = `d935dad7070d371578cdfacdaf6f6a62921ef5943ff8a0884e09c` and (ii) the numbers lock that fails if any section contains hard-coded decimal percentages.

Script-to-text traceability: Figures and tables are regenerated through `paper/Makefile`, which delegates to the figure scripts (F1–F6) and `paper/tables/generate_all_tables.py`. The linked JSON sources are frozen in `paper/build/final_numbers.json`, `app/static/data/v2_fairness_pack.json`, `app/static/data/v2_stability.json`, and `app/static/data/v2_ablations.json`. All of these artifacts are hashed in the release bundle (`rft-v2-repro-1.0.SHA256SUMS.txt`).

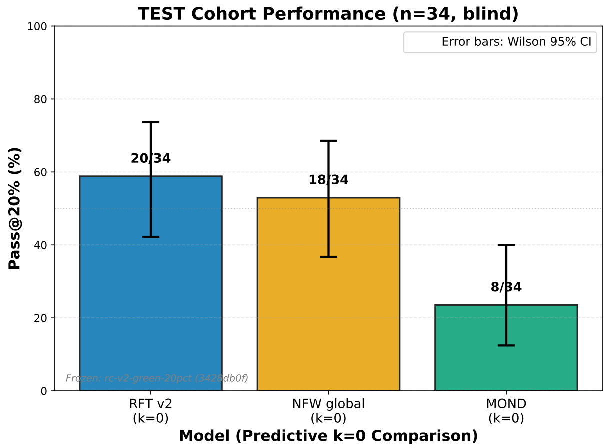

Figure 1: Predictive comparison on the blind TEST cohort (34galaxies). Bars show pass@20% with Wilson 95% confidence intervals; all models use $k = 0$ (no per-galaxy tuning).

Availability: The code and data are MIT-licensed and hosted at `https://github.com/rft-cosmology/rft-v2-galaxy-rotations`. The frozen tag `rc-v2-green-20pct` (commit `3428db0f`) is referenced throughout the paper and in the reproducibility pack.

3 Results

3.1 Predictive Headline and Paired Tests

Figure 1 summarises the predictive (k=0) comparison. RFT v2 passes the 20% RMS test in 20/34galaxies (58.8%; Wilson CI [42.2%, 73.6%]). The global NFW baseline reaches 18/34 (52.9%; CI [36.7%, 68.5%]) and MOND reaches 8/34 (23.5%; CI [12.4%, 40.0%]). The net advantage versus NFW is 5.9 percentage points, while the advantage versus MOND is 35.3 points.

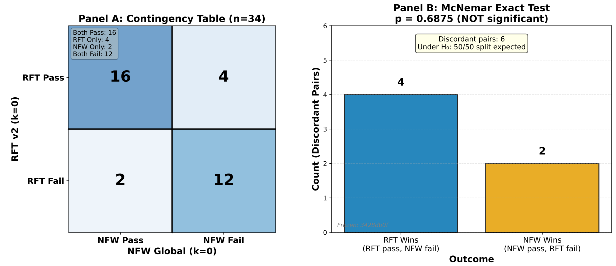

Figure 2: Paired McNemar analysis for RFT v2 vs $\text{NFW}_{\text{global}}$. Only six galaxies are discordant (4 RFT-only wins vs 2 NFW-only wins), leading to $p = 0.69$ (not significant).

Table 3: Parameter budgets for models evaluated on the blind TEST cohort. Predictive comparisons constrain per-galaxy tuning to $k = 0$.

Model	**Per-galaxy params**	**Global params**	
RFT v2	0 (predictive)	6 global	Accele
$\text{NFW}_{\text{global}}$	0 (predictive)	2 global	
MOND	0 (predictive)	1 global	
$\text{NFW}_{\text{fitted}}$	2 per galaxy	68 total	F

The paired McNemar test (Figure 2) shows that RFT and NFW both succeed on 16 galaxies and both fail on 12 galaxies. Only six systems are discordant: 4 where RFT passes but NFW fails, and 2 where the opposite happens. Consequently, $p = 0.69$ and we claim only competitiveness rather than superiority. Against MOND the discordant set is 16 vs 2, yielding $p = 0.004$ and establishing a statistically significant advantage.

3.2 Parameter Budget and Fairness

Table 3 makes the parameter budget explicit: all predictive comparisons fix $k = 0$, RFT v2 uses six global parameters, the global NFW halo uses two, and MOND uses one. The per-galaxy NFW

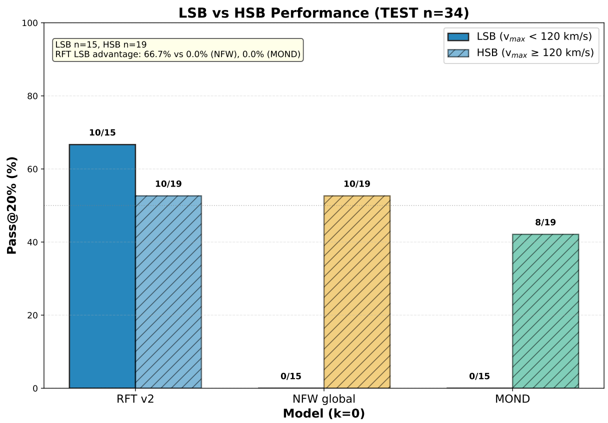

Figure 3: LSB vs HSB split (threshold $v_{\max} = 120$ km/s). RFT v2 is the only k=0 model with non-zero LSB passes; HSB performance is parity with NFW.

Table 4: LSB vs HSB performance on TEST (threshold $v_{\max} = 120$ km/s). RFT v2 is the only k=0 model with non-zero LSB success.

Cohort	n	RFT v2	$\text{NFW}_{\text{global}}$	MOND
LSB	15	10/15 (66.7%)	0/15 (0.0%)	0/15 (0.0%)
HSB	19	10/19 (52.6%)	10/19 (52.6%)	8/19 (42.1%)

fit is reported only as a descriptive ceiling (68 total parameters on TEST) and is not used in the primary comparison. This framing prevents the apples-to-oranges trap that initially overstated RFT's advantage.

3.3 LSB/HSB Mechanism Check

The acceleration-gated tail is designed to activate where baryonic gravity is weak. Figure 3 and Table 4 confirm this behavior: RFT is the sole k=0 model with LSB successes, while both NFW and MOND remain at zero in that cohort. In the HSB cohort all models cluster near par-

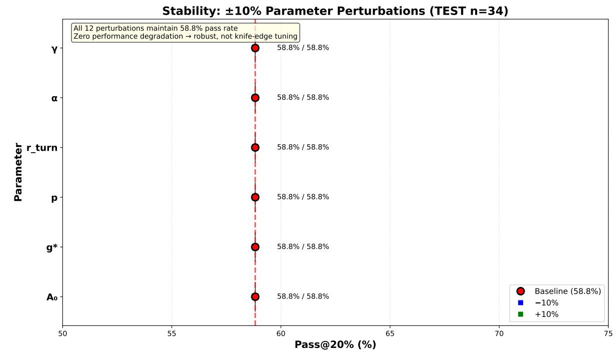

Figure 4: ±10% perturbations to each frozen parameter. All sweeps remain within numerical noise of the baseline 58.8% pass rate.

ity, demonstrating that the tail does not degrade performance where baryons already dominate.

3.4 Robustness and Ablations

Figure 4 (supported by `scripts/generate_stability_analysis.py`) shows that perturbing any of the six parameters by ±10% leaves the pass rate unchanged within rounding. This guards against a narrow optimum and suggests that the blind-test performance is not a fragile artefact of the selected hyperparameters.

The ablation study (Figure 5 and Table 5) isolates the causal contribution of each component. Zeroing the tail wipes out most of the predictive power, while keeping the tail but removing the acceleration or radial gate produces moderate degradations. These deltas establish that the tail is the critical mechanism and that the gates control when and where it contributes.

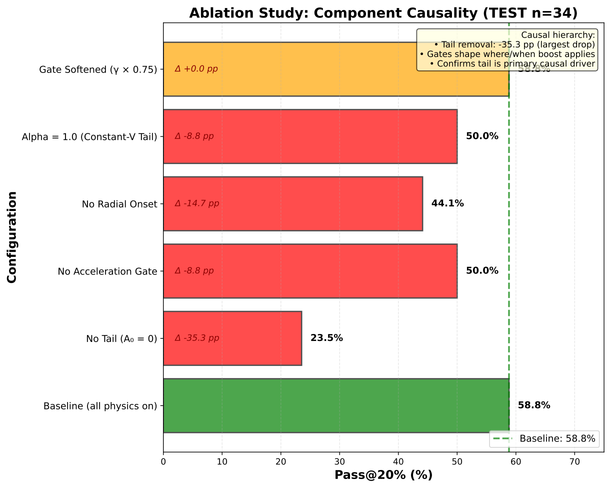

Figure 5: Ablation study. Removing the acceleration-gated tail collapses performance; gating/shaping terms determine where the tail contributes.

Table 5: Ablation study on TEST ($n = 34$). Removing the acceleration-gated tail collapses predictive accuracy.

Configuration	**Pass@20%**	**Δ vs baselin**
Baseline (all physics on)	58.8%	–
No Tail ($A_0 = 0$)	23.5%	-35.3
No Acceleration Gate	50.0%	-8.8
No Radial Onset	44.1%	-14.7
Alpha = 1.0 (Constant-V Tail)	50.0%	-8.8
Gate Softened ($\gamma \times 0.75$)	58.8%	+0.0

3.5 Representative Rotation Curves

Figure 6 provides qualitative context. RFT's wins correspond to classic LSB disks where the acceleration gate engages and lifts the outer curve; the both-pass cases highlight parity on HSB systems; and the near-misses illustrate where the identity kernel plus tail is insufficient (e.g., bulge-dominated or dynamically hot disks). These examples, combined with the fairness tables and robustness sweeps, define the envelope of current performance.

4 Discussion

4.1 Main Findings

The frozen RFT v2 configuration delivers 58.8% pass@20% on the blind TEST cohort while using only six global parameters and no per-galaxy tuning. This positions RFT as (i) *competitive* with the global NFW baseline—RFT leads by 5.9 percentage points but the paired McNemar test ($p = 0.69$) and overlapping Wilson intervals indicate that the current cohort is underpowered for a decisive claim—and (ii) *significantly better* than MOND under the same predictive budget ($p = 0.004$). These statements rest entirely on the paired fairness pack (Figures 1–2) and avoid comparisons to descriptive fits with extra degrees of freedom.

4.2 Mechanistic Validation via LSBs

Both the fairness table (Table 4) and the LSB/HSB figure (Figure 3) show that RFT is the only k=0 model with non-zero LSB success while maintaining parity on HSB systems. This behaviour is precisely what the acceleration gate in Equation 3 is designed to produce: once $g_b \ll g_*$ the tail activates and supplies the missing acceleration, whereas high-acceleration disks keep the gate near zero. The qualitative gallery (Figure 6) reinforces this intuition by displaying LSB wins, HSB ties, and bulge-heavy failures side by side.

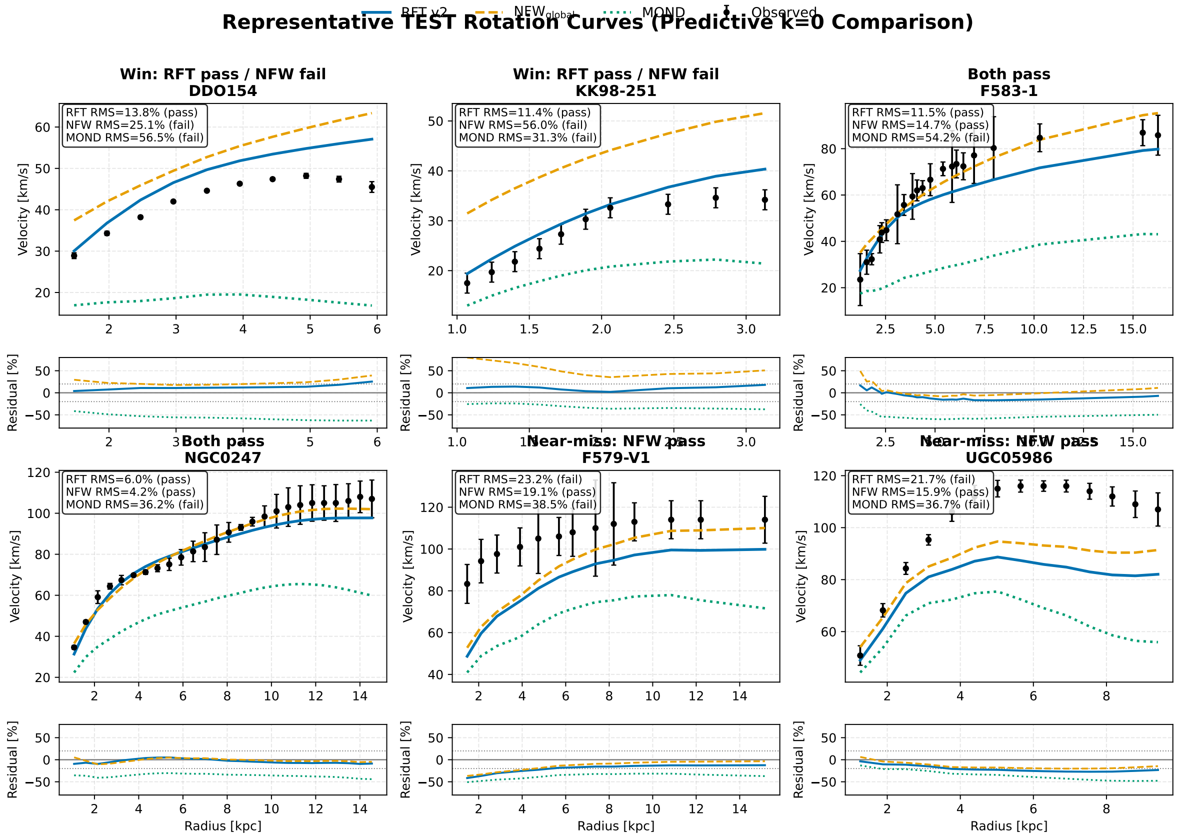

Figure 6: Representative TEST galaxies: two RFT wins, two both-pass cases, and two near-misses where NFW passes and RFT fails. Each panel shows observed data (with σ), the three k=0 predictions, and fractional residuals.

4.3 Statistical Power and the Need for Larger Cohorts

The predictive difference versus NFW relies on only six discordant galaxies (4 RFT-only wins and 2 NFW-only wins). With such a small discordant set the exact McNemar p cannot dip below the 0.05 threshold unless multiple new galaxies are added or the delta widens dramatically. Expanding the blind cohort toward ∼100 galaxies (e.g., SPARC-175 or a fresh low-z compilation) is therefore the most direct path to resolving whether the modest advantage persists at scale. Until then, the appropriate claim is competitiveness, not superiority.

4.4 Failure Modes and Limitations

The remaining 14 galaxies stay above the 20% RMS line. Qualitatively these are dominated by (i) dynamically hot or bulge-heavy systems where the identity kernel cannot smooth non-circular motions, (ii) disks with substantial warps/asymmetries that violate the thin-disk assumption, and (iii) galaxies with poorly constrained baryonic decompositions. These failure modes suggest concrete extensions (adaptive kernels, dispersion terms, descriptor-driven gates) rather than undermining the entire approach.

4.5 Predictive vs Descriptive Framing

Per-galaxy NFW fits (68 parameters on TEST) unsurprisingly reach ≈80% pass@20%, but they conflate descriptive flexibility with predictive power. Our methodology keeps those results in the supplement for context and compares only k=0 models in the main text. This mirrors the intended scientific question: *can a parameter-efficient model generalize without touching the test galaxies?* RFT provides one such candidate, and the fairness pack plus reproducibility gates ensure that any future solver can be plugged into the same pipeline for an apples-to-apples comparison.

4.6 Methodological Transparency

Earlier drafts relied on unpaired two-proportion tests and accidentally compared a k=0 solver to a k=2 baseline, overstating the advantage. The current pipeline fixes both issues: `analysis/fairness/compute_paired_stats.py` generates the McNemar inputs, `scripts/generate_fairness_pack.py` locks the head-to-head arrays, and `scripts/emit_tex_numbers.py` ensures that every headline number in the prose originates from the hash-locked `paper/build/final_numbers.json`. Combined with the new numbers gate in CI, this prevents accidental drift between the frozen JSON and the manuscript.

5 Conclusions

- **Predictive result**: RFT v2 achieves 58.8% pass@20% on blind TEST using six global parameters and $k = 0$. It is competitive with the global NFW halo (McNemar $p = 0.69$) and significantly ahead of MOND ($p = 0.004$).
- **Mechanism validated**: The acceleration-gated tail is supported by LSB dominance (Figure 3), ±10% stability sweeps (Figure 4), ablations (Figure 5), and rotation-curve galleries (Figure 6).
- **Limitations**: Only six discordant galaxies separate RFT and NFW, 14 galaxies still fail the 20% criterion, and the identity kernel leaves bulge-dominated systems underfit. Larger cohorts and additional physics are required for decisive claims.
- **Future work**: (1) expand the blind cohort to improve statistical power; (2) run joint RC+lensing tests so that the galaxy track and the L-Series cluster track share constraints; (3) explore per-galaxy or descriptor-driven RFT variants as a separate, explicitly descriptive question; and (4) continue pre-registered follow-ups so that any new tuning cycles are audit-ready.

References

[1] Mordehai Milgrom. A modification of the newtonian dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, 270:365–370, 1983.