Note: In this document, α refers exclusively to the fine-structure constant (α-in) used as input for calibration. The symbols β and γ refer only to standard Post-Newtonian (PPN) parameters relevant to external validation protocols. No structural potential terms {α_V, β, γ} appear in this document.

Limitations & Validation Plan — Structural Field Theory (SFT)

Purpose. This note consolidates current limitations of SFT and the corresponding validation plan. It is written as a self-contained section to be included in the manuscript bundle.

Language guardrail (recommended): prefix validation items with (C) calibrated/CI-synthetic/methodology vs (P) prediction/REAL-ready inputs required. Use (C) for pipeline auditability and (P) only when REAL solver outputs are in play.

1) Experimental validation is pending

Concern. While SFT makes concrete predictions, direct tests (e.g., sub-millimeter torsion measurements; CMB/FRB birefringence) are not yet completed.

Response. We treat this as an opportunity and commit to a short, testable roadmap:

- **(I) Static Coulomb calibration (α-in): use α_ref as INPUT for calibration; report only consistency with α_ref. Any α-out estimation, if executed, lives in a pre-registered appendix with a PASS/FAIL threshold (e.g., τ = 1%) and does not affect RC validity.**
- **Astrophysical cross-checks.** Recast SFT's birefringence/dispersion predictions against existing CMB TB/EB bounds and FRB/GRB polarization catalogs; declare pass/fail criteria ex ante.
- **Quantitative acceptance bands.** For each test we publish numerical thresholds (energy conservation ΔE/E, phase-speed dispersion |v_phase−c|/c, PPN γ,β), so that third parties can reproduce and decide independently.

2) Scope relative to the Standard Model is limited (for now)

Concern. Weak interactions, flavor, and the full hadronic spectrum are not modeled yet.

Response. SFT's present claims are deliberately scoped to the EM+gravity sector and to the emergent spin-½ route. The extension path is explicit:

- **Dirac/Wilson appendix completion.** Add numerical checks (no doublers, γ5-hermiticity, convergence) with tolerances.
- **Internal-texture routes to non-Abelian structure.** "Route B" remains historical/optional, not required for the mainline; we will report negative results as such.
- **Scope discipline.** Until the above milestones are met, we avoid over-reach and label out-of-scope items as future work.

3) Initial calibration may look like tuning

Concern. The three‑step calibration of q*, ħ*, ε*, μ* could be perceived as parameter‑tuning.

Response. We operate a single-pass calibration pipeline (already stated across the corpus): calibrate emergent scales once from {α_em, c} via a static Coulomb test, freeze them, and then run all demonstrations without per-observable retuning. To make this audit-proof we will:

- **Leave-one-out stress test.** Drop α_em from calibration → predict it back; report residual.
- **Traceability table.** Each figure/number lists precisely which frozen scales it consumes. This turns "tuning" into a reproducible metrology step with cross-validation, rather than free fitting.

Mini-example (illustrative, no numbers):

Item	Consumes frozen scales	Retuning allowed?
Static Coulomb calibration	{q*, ħ*, ε*, μ*} frozen from {α_em, c}	No (single-pass)
Demo observable X	Frozen scales only (no per-observable fit)	No
α-out appendix	Separate preregistered gate_id family	No post-hoc

4) Discrete vs. continuum tension

Concern. Even with a clean continuum limit, discreteness can jar with traditional QFT/GR.

Response. Our mainline Lagrangian is Lorentz-invariant in the continuum; discretization induces O((a k)^2) corrections that we measure and report (slope ξ/2 in the dispersion-fit). We provide: (I) a convergence scan in a (2^{nd}-order as claimed); (II) an explicit gauge-fixed Noether construction showing that the Maxwell sector carries no extra DOF (A is a functional of S + lattice operators). This reframes discreteness as a numerical regularization with controlled systematics, not as an alternative kinematics.

Audit anchor: the convergence/EOC scan is operationalized in Doc 6 (EOC/CI) and reported via the standard SCAN/REGION/REPORT artifacts where applicable. Reviewers should look for the reported O((ak)^2) slopes and resolution tags in those artifacts.

Reviewer Checklist (one page)

1. **Build & Repro:** Docker image builds; pytest passes.
2. **Calibration:** Run once, record q*, ħ*, ε*, μ*; no retuning afterward.
3. Numerics: Acceptance bands (preregistered targets): report ΔE/E ≤ 1e−3; continuity residual ≤ 1e−4; dispersion error ≤ 1%.

4. **PPN:** State convention (signature −,+,+,+, S = −U), show $\gamma = 1$ (linear), and $\beta = 1 + c\_\beta \lambda_4 + O(\lambda_4^2)$ with the DSM protocol to extract c_β.
5. **EM emergence:** Cite the Noether→Maxwell pipeline and uniqueness/gauge-fixing note (no extra DOF).
6. **Scope:** Label weak/flavor/hadrons as future work; list milestones and negative-result policy.
7. **Experiments:** Define pass/fail bands for (I) sub-mm torsion, (II) CMB/FRB birefringence, (III) lab dispersion.

Cross-links. See Integrated Technical — Unified Notation & Units (PPN convention), Appendix B (Emergent EM from S), and the Simulation Supplement (Reproducibility summary; defaults; checklist).

Reviewer Checklist — language guardrail: use labels (C) calibrated / (P) prediction; avoid "predicts/reproduces α" wording in the RC body.

APPENDIX A

External Break Test #01 — Field Rescaling (Invariance vs Scaling Law)

Audit protocol (Doc9/Doc15-style) to detect hidden scale dependence and post-hoc tuning

Status: CI / Methodology. Break test intended for external verification.

Non-claim: Does not assume S→λS is a physical symmetry. It evaluates (i) representation invariances when applicable and (ii) expected scaling laws when not.

1. Objective

Detect whether the pipeline (estimators, normalizations, gates, and reports) contains hidden scale dependence or compensatory calibration. This test attempts to break the system with a simple field rescaling to reveal: (a) leakage through normalization, (b) adaptive gates, and (c) undeclared sensitivity to amplitude.

2. Definitions

Base field: $S(x)$ generated or imported for a test case with a fixed seed.

Rescaling: $S\_\lambda(x) = \lambda \cdot S(x)$, with $\lambda \in \{0.5, 1, 2\}$ (minimum).

The λ set used for evaluation MUST be declared in DECISION_RULE.json (and hashed) before any run; changing λ post-hoc invalidates the run.

DECISION_RULE.json: preregistered definition of operators (stencil/BC), normalization, metrics, and thresholds. Must be hashed (decision_rule_sha256) and referenced by gate_id.

3. Preconditions (prereg / anti-post-hoc)

- Freeze DECISION_RULE.json before running (gate_id + decision_rule_sha256).
- Deterministic seeds (seed) and version/commit recorded in the report.
- Same grid, same BCs, same dt/CFL (if applicable), same configuration except for the rescaling λ.
- JSON sanitization: NaN/Inf → null before hashing artifacts.
- No threshold/window adjustments after seeing results (post-hoc loosening is forbidden).

4. Two complementary tests

Test A — Representation invariance (if applicable)

Use this when field rescaling is a representational redundancy (e.g., S defined up to a multiplicative factor), or when a preregistered parameter transform exists that leaves the model equivalent.

If Test A is used, the exact remap $p \to p(\lambda)$ (and any unit/normalization conventions) MUST be stated in DECISION_RULE.json and included in the manifest via decision_rule_sha256.

Expectation: invariance of outputs explicitly declared as “representation-level” under $S \to \lambda S$.

The list of representation-level outputs MUST be declared in DECISION_RULE.json (and hashed), to prevent post-hoc relabeling.

Test B — Physical scaling law (if A does NOT apply)

Use this when $S \to \lambda S$ is not a symmetry. In this case invariance is NOT required; instead, require a preregistered expected scaling law (e.g., energies ~ λ^2, amplitudes/gradients ~ λ, etc., depending on the model).

For each observable y, DECISION_RULE.json must specify y_expected(λ) (functional form or exponent), the evaluation domain/window, and any masking/exclusions used.

Expectation: metrics follow the declared scaling law; otherwise FAIL (or “unknown”).

5. Procedure (step-by-step)

1. Choose a base case (S, configuration, seed) and produce/obtain the input artifacts.

2. Create three variants: $\lambda = 0.5, 1, 2$ (or 1/3, 1, 3 if you prefer larger separation).

3. For each λ: run the full pipeline WITHOUT changing anything else.

4. Save the standard artifacts (SCAN/REGION/REPORT if applicable) and the SHA-256 manifest.

5. Compare metrics across λ and evaluate gates according to Test A or Test B.

6. Expected artifacts

Note (naming): artifact filenames may be prefixed with BREAKTEST01_* for cross-pack reuse, provided the internal schema fields (gate_id, decision_rule_sha256, inputs_sha256) are unchanged.

Minimum (per λ):

- ALPHA_*_REPORT.json (includes: suite_version, gate_id, decision_rule_ref, decision_rule_sha256, provenance, GLOBAL_PASS).
- Optional: COMPATIBILITY_SCAN_*.json and EXISTENCE_REGION_*.json (if the pack uses them).
- MANIFEST_SHA256.txt (includes hashes for the report + DECISION_RULE.json + inputs).
- Optional plots: sensitivity_sweep.png, confusion_matrix.png (non-authoritative).

7. Recommended metrics and gates

Record (per λ):

- alpha_like (if applicable) and/or alpha_out (if applicable), with relative drift vs $\lambda=1$.
- Intermediate metrics feeding alpha (e.g., invariants, energy, gradients, fits).
- Scale-sensitivity score per observable: $s = \max(|\Delta \log y|) / |\Delta \log \lambda|$ (report per y).
- Null/adversarial indicators (if the pack has them): FP_rate, F1, etc.

Suggested gates (examples; must be preregistered):

- Test A: $|\Delta\, y_{rep}| / (|y_{rep}|+\varepsilon) \leq \tau_{rep}$ for representation-level outputs.

Example (non-normative): Suggested CI defaults (must still be pre-registered): $\tau_{rep} = 1e{-3}$ and $\tau_{scale} = 1e{-2}$, unless numerical analysis justifies tighter/looser values.

Non-normative examples only. Thresholds must be preregistered per pack (gate_id) and justified or inherited from pack-level numerical analysis.

- Test B: $|\log y - \log y_{expected}(\lambda)| \leq \tau_{scale}$ for observables with a declared scaling law.
- Anti-compensation: if alpha_out remains “suspiciously constant” while base metrics change strongly, mark as FAIL or “needs review”.

8. Quick interpretation (what PASS/FAIL means)

- PASS (Test A): representation-level outputs invariant within gates; manifests and hashes consistent.
- PASS (Test B): outputs follow the preregistered scaling law; deviations small and explainable.
- FAIL: undeclared scale dependence, adaptive gates, post-hoc window changes, or artifact/hash inconsistency.
- INCONCLUSIVE: neither symmetry nor preregistered scaling law; declare “unknown” and preregister expectations before repeating.

9. Where this fits in the corpus

This test is recommended as “External Break Test #01” within the validation plan (e.g., Doc 8 or as an appendix to Doc 15). It is especially useful to audit that normalizations, thresholds, and decision rules do not introduce implicit calibration.

Structural Field Theory. Complete theory, documents and runners at: https://github.com/Xaquer69/sft-theory-and-runners

https://doi.org/10.5281/zenodo.17608314