

Skeptical review: GPU-Accelerated Particle-Mesh Cosmological Simulations with NVIDIA Warp: Performance and Accuracy Validation

Summary

This manuscript presents a GPU-accelerated cosmological Particle–Mesh (PM) N-body code implemented in NVIDIA Warp. The authors evolve 10 realizations of a $(1000 \text{ Mpc}/h)^3$ volume with 512^3 particles on a 512^3 mesh from $z = 127$ (2LPT initial conditions) to $z = 0$ using a leapfrog integrator with 200 steps (Secs. 2.1–2.4), and validate primarily via the matter power spectrum. The reported runtime ($\sim 20 \text{ s}$ per realization on a single GPU; Sec. 3.1) is potentially very impactful for producing large ensembles for covariance and large-scale statistics. However, the current version has (i) an important inconsistency about the benchmark used for validation (Quijote vs CAMB HaloFit across the Abstract, Sec. 2.4, and Sec. 3.2), (ii) insufficient numerical and implementation detail for the PM solver, integration, and $P(k)$ measurement to be reproducible or to interpret the scale-dependent suppression, and (iii) performance claims that are difficult to assess without concrete hardware/software specification and a measured CPU baseline. The physical validation is also narrow (single box/resolution/cosmology and mostly $z = 0$ $P(k)$), and the $\sim 4\%$ large-scale offset (ratio ~ 0.96) relative to the stated reference should be diagnosed (normalization/growth/measurement pipeline) before making broader claims (Secs. 3.2–4). Addressing these points would substantially strengthen both the technical credibility and the scientific usefulness of the work.

Strengths

- The paper is well motivated by the need for large ensembles of cosmological simulations for cosmic-variance control and covariance estimation, and it clearly positions GPU PM methods as a pragmatic approach (Sec. 1).
- The overall pipeline—2LPT initial conditions, CIC mass assignment, FFT-based Poisson solve, leapfrog time integration, and power-spectrum estimation with CIC deconvolution and shot-noise subtraction—is standard and appropriate for the stated large-scale goals (Secs. 2.2–2.4).
- The reported single-GPU throughput ($\sim 20 \text{ s}$ per 512^3 -particle realization with 200 steps) is potentially very useful for rapid generation of ensembles, and the manuscript makes a clear case for why such speed matters (Sec. 3.1).
- Ensemble-based reporting (10 realizations) is a good choice to illustrate cosmic variance and run-to-run scatter, rather than relying on a single realization (Sec. 3.2).
- Figures use a clear mean-and-scatter presentation (mean curve plus shaded band), and the two-panel $P(k)$ /ratio layout is an effective diagnostic format (Fig. 2 / Sec. 3.2).

Major issues

1. **Validation benchmark is inconsistent and currently unclear (Quijote vs CAMB HaloFit).** The Abstract and Sec. 2.4 refer to comparison against the “high-fidelity Quijote simulation suite,” while Sec. 3.2 and the plotted comparisons describe a nonlinear CAMB HaloFit prediction, with no explicit Quijote comparison shown. These are materially different references (simulation vs fitting formula), and the inconsistency undermines the accuracy claims and interpretation of the reported percent-level agreement (Abstract; Secs. 2.4, 3.2; Fig. 1/2 captions; Sec. 4).

Recommendation: Make the validation reference consistent throughout. If HaloFit is the main reference, revise the Abstract, Sec. 2.4, and Sec. 4 to say so, and explicitly acknowledge limitations of treating HaloFit as “truth” at the few-percent level (especially beyond linear scales). If Quijote is intended as the benchmark, add a dedicated comparison in Sec. 3.2 (ratio to Quijote $P(k)$ with quantitative metrics over a stated k -range), and document exactly which Quijote runs/outputs are used (resolution, method, box size, redshifts), plus how $P(k)$ is measured/matched (mass assignment, deconvolution, binning, shot noise) so the comparison is like-for-like.

2. **Core PM gravity solve and integration details are underspecified, preventing reproducibility and obscuring the origin of scale-dependent power suppression.** In Sec. 2.3 and Eq. (3), the Poisson solve is written as $\hat{\phi}(k) = -4\pi G \hat{\rho}(k)/k^2$ but the comoving cosmological form is not fully specified (physical vs comoving density; use of overdensity δ ; any factors of a and $\bar{\rho}$; treatment of the $k = 0$ mode). Additional key implementation choices are missing: continuous k^2 vs discrete lattice Green’s function, whether CIC deconvolution is applied in the force computation (not just in measuring $P(k)$), how forces are differenced (stencil/order), boundary conditions (presumably periodic), and whether any additional softening exists beyond the mesh (Sec. 2.3).

Recommendation: Expand Sec. 2.3 (or add an Appendix referenced from Sec. 2.3) with the exact equations and discrete operators implemented: (i) write the comoving Poisson equation being solved and define the source field (ρ , $\rho - \bar{\rho}$, or δ), including any a -dependent prefactors; (ii) state explicitly how the $k = 0$ mode is handled; (iii) specify the Fourier-space Green’s function (continuous $-1/k^2$ vs discrete Laplacian/lattice Green’s function) and any force-kernel filtering; (iv) state whether and how CIC is deconvolved in the force computation and how forces are interpolated back to particles; (v) specify the finite-difference stencil used to compute $\nabla\phi$ on the grid and the assumed periodic boundary conditions.

3. **Time integration choices are not justified and no convergence evidence is provided.** Sec. 2.3 states 200 timesteps uniformly spaced in scale factor a and uses a leapfrog method, but does not specify the leapfrog variant (KDK vs DKD), the exact update variables (peculiar velocity vs canonical momentum), how Hubble drag and $aH(a)$ enter the updates, or why 200 steps is sufficient for the claimed large-scale accuracy. Without a basic timestep convergence test it is difficult to separate integration error from force-resolution error in Sec. 3.2.

Recommendation: In Sec. 2.3, provide the explicit leapfrog update equations, state KDK vs DKD, and clarify whether you integrate velocities or momenta (and how Hubble drag is included). Add a minimal convergence check in Sec. 3.2 (e.g., $N_{\text{steps}} = 100/200/400$) reporting fractional differences in $P(k)$ at $z = 0$ (and ideally also at an intermediate redshift) over the k -ranges used for claims (e.g., $k < 0.03 h/\text{Mpc}$). If additional runs are not feasible, soften/qualify accuracy claims and cite prior PM timestep guidance consistent with your setup.

4. **Performance claims are difficult to interpret without a measured baseline and complete timing methodology.** Sec. 3.1 (and the Abstract) compare ~ 20 s on GPU to an “estimated” ~ 5 hours on CPU without specifying CPU code, hardware, threading, FFT library, or measurement procedure. The GPU timing itself is not fully defined (does it include IC generation, FFT plan creation, I/O, and $P(k)$ measurement, or only the time-stepping loop?). Without this, the implied orders-of-magnitude speedup could be misleading and is not reproducible (Sec. 3.1).

Recommendation: In Sec. 3.1, provide a clear and reproducible timing methodology: list GPU model + memory, host CPU/RAM, OS, CUDA/driver versions, Warp version, FFT backend (e.g., cuFFT), and precision (float32/float64) for particles/FFTs. Define exactly what is included in the reported ~ 20 s. Add a measured CPU baseline for the same algorithm (or a standard PM code) on specified CPU hardware with core count, compiler flags, threading (OpenMP/MPI), and FFT library; if not feasible, clearly label the 5-hour value as a rough estimate, explain how it was obtained, and tone down language in the Abstract/Sec. 4 accordingly. A simple kernel-level breakdown (CIC deposit, FFT Poisson, force interp, drift/kick) would further strengthen Sec. 3.1.

5. **Physical setup and initial-condition details are incomplete, limiting reproducibility and potentially affecting the reported large-scale offset.** Across Secs. 2.1–2.2, the manuscript does not fully specify: the exact cosmological parameters (e.g., Quijote fiducial values), the tool and settings used to compute the linear power spectrum at $z = 127$ (CAMB/CLASS version/config), how growth factors f_1 and f_2 are obtained, random seed/phase generation for the 10-realization ensemble, and whether phases

are matched to any external benchmark (if Quijote comparison is intended). These omissions make it hard to diagnose the ~ 0.96 large-scale ratio noted in Sec. 3.2 (possible normalization/growth mismatch vs measurement pipeline mismatch).

Recommendation: Add a self-contained table (Sec. 2.1) listing the cosmological parameters used and explicitly cite the source (Quijote fiducial, if applicable). In Sec. 2.2, state the linear-theory code + version + key settings (transfer function, normalization, massive neutrinos if any), how f_1/f_2 are computed, the grid used for the Gaussian random field and displacement fields, and the RNG/seed strategy for independent realizations. If comparing to Quijote, clarify whether initial phases are matched; if not, avoid implying a realization-by-realization comparison. In Sec. 3.2, comment explicitly on possible causes of the $\sim 4\%$ large-scale offset and add a sanity check (e.g., compare early-time $P(k)$ to linear theory at high z , or verify that σ_8 of the ICs matches the reference).

6. **Power-spectrum measurement and uncertainty treatment lack key specifications (binning, Fourier conventions, aliasing control, ratio definition), and this affects both reproducibility and interpretation of “few-percent” agreement. Sec. 2.4 does not fully specify the Fourier normalization convention, the exact CIC window used for deconvolution, k -bin edges/spacing, treatment near Nyquist, or whether any anti-aliasing/interlacing is used. Sec. 3.2/figures do not unambiguously state whether ratios are computed as $\langle P \rangle / P_{\text{ref}}$ or $\langle P / P_{\text{ref}} \rangle$, and the shaded band risks being misread as error-on-the-mean rather than scatter across volumes (Sec. 2.4; Sec. 3.2; Fig. 2).**

Recommendation: In Sec. 2.4, explicitly define $\delta(x)$, your FFT normalization, and write $P(k)$ consistently under that convention so the shot-noise subtraction term is unambiguous. Specify k -binning (bin edges, linear/log spacing, number of bins, mode weighting), Nyquist frequency, and whether modes near Nyquist are excluded. Provide the explicit CIC window $W_{\text{CIC}}(\mathbf{k})$ used for deconvolution and describe any regularization near zeros. State whether you use interlacing or another aliasing-control technique; if not, discuss expected aliasing impact and restrict/flag affected k -ranges. In Sec. 3.2 and the Fig. 2 caption, state precisely how the mean/ratio and the shaded band are computed (standard deviation across 10 realizations vs standard error on the mean), and consider showing both if you interpret mean offsets at the percent level.

7. **Validation scope is narrow relative to the breadth of claims. Results focus on a single box size, single particle/mesh resolution, a single cosmology, and primarily $z = 0$ $P(k)$ (Secs. 2.1, 3.2), while Sec. 4 suggests broad appli-**

capability for BAO studies and covariance estimation. Without at least minimal checks across redshift and/or resolution, it is unclear how robust the conclusions are for typical analysis workflows.

Recommendation: Either broaden validation or narrow/qualify claims. If feasible, add in Sec. 3.2: (i) $P(k)$ at one or two intermediate redshifts (e.g., $z = 1, 0.5$), and/or (ii) a simple resolution/mesh test (e.g., 256^3 vs 512^3 mesh at fixed particles, or vice versa) to demonstrate expected convergence trends. If additional runs are not feasible, revise Sec. 4 to explicitly limit claims to the demonstrated regime (large scales, this specific setup) and clarify which applications are plausible (e.g., large-scale-only statistics) versus not yet validated (halo-scale/non-Gaussian/covariances requiring accurate mode coupling).

Minor issues

1. Figure 2 (and related text in Sec. 3.2) is missing or unclear metadata: axis units for $P(k)$, explicit redshift, denominator for the ratio panel, number of realizations in the legend, and visual markers for the k -thresholds discussed in the caption/text (e.g., $0.03, 0.05, 0.1 h/\text{Mpc}$). The current rendering size risks illegible text/bands in print.

Recommendation: Update Fig. 2 to include $P(k)$ units (e.g., $h^{-3} \text{Mpc}^3$), explicitly label $z = 0$, specify the ratio denominator in the panel label/legend, and include “ $N = 10$ ” in the legend. Add vertical guide lines/annotations at $k = 0.03/0.05/0.1 h/\text{Mpc}$ so the stated regimes are visually traceable. Increase font sizes/panel height and provide a vector figure (PDF/SVG) to ensure readability.

2. Limitations are not stated prominently enough given the paper’s positioning. The manuscript notes PM small-scale suppression (Sec. 3.2) but does not clearly enumerate limitations (PM-only, grid force resolution, no short-range correction, single-species DM-only, limited validation) while Sec. 4 suggests broad use for BAO and covariance analyses.

Recommendation: Add a concise limitations paragraph in Sec. 4 explicitly listing known constraints and the validated regime, and briefly outline straightforward extensions (e.g., P3M/TreePM short-range correction, higher mesh, multi-species, additional summary statistics). This will help prevent overinterpretation.

3. Hardware/software environment and numerical precision are not fully specified (Secs. 2.1, 3.1), yet they strongly affect both performance and accuracy (especially over 200 steps and large FFT workloads).

Recommendation: Add a short “Computational environment” paragraph (Sec. 2.1 or 3.1): GPU model/memory, host CPU/RAM, OS, CUDA/driver, Warp version, FFT backend, and whether particle state/forces/FFTs are float32 or float64. Briefly comment on precision/accuracy trade-offs for the reported setup.

4. Ambiguous notation “5123” appears where 512^3 is intended (Secs. 1, 2.1, 2.3, 3.1, 4), which can confuse readers and makes it harder to sanity-check derived quantities (particle count, mesh size, \bar{n} , shot noise).

Recommendation: Replace all occurrences with unambiguous LaTeX (512^3) and, where helpful, also provide the integer total (134, 217, 728) once for clarity.

5. Shot-noise subtraction and Fourier conventions are not fully pinned down (Sec. 2.4). Stating the shot-noise correction as exactly $1/\bar{n}$ without defining δ and the Fourier normalization makes it hard to verify correctness and to reproduce the exact $P(\mathbf{k})$ pipeline.

Recommendation: In Sec. 2.4, define δ in terms of number density (or mass density) and state your FFT normalization. Then write the corresponding shot-noise term under the same convention (including any volume factors if applicable) and clarify whether shot noise is subtracted before/after CIC deconvolution.

6. 2LPT implementation details remain somewhat conceptual (Sec. 2.2): whether an external 2LPT package is used, whether displacement fields are computed on the same 512^3 grid as the PM mesh, whether any smoothing/anti-aliasing is applied, and how velocities are constructed numerically from f_1/f_2 .

Recommendation: In Sec. 2.2, state the code/library used (or that it is custom), the grid resolution for displacement fields, how f_1/f_2 are computed (analytic vs numerical growth), and whether any smoothing/anti-aliasing is used when generating and transforming fields.

Very minor issues

1. Typographical/LaTeX issues: line-break artifacts splitting words (e.g., “gen erate”, “sim ulation”), inconsistent unit spacing (Mpc/h vs Mpc/h ; h/Mpc vs h/Mpc), and inconsistent inequality symbols in figure text/captions (Secs. 1, 2.1, 3.2, 4; Fig. 2).

Recommendation: Proofread to remove hyphenation/line-break artifacts and standardize unit formatting and inequality notation across the manuscript and figure captions.

2. Equation presentation and variable naming could be clearer (Secs. 2.2–2.4). Some equations would benefit from displayed formatting and explicit variable definitions; the use of \vec{q} for displaced/Eulerian positions may confuse readers given \vec{x} is already used as the Lagrangian coordinate.

Recommendation: Typeset Eqs. (1)–(4) as displayed equations with consistent definitions immediately nearby, and add a brief clarification of \vec{q} (or rename) to avoid convention clashes.

3. The CIC deconvolution is described qualitatively but the explicit CIC window function is not written (Sec. 2.4), making the exact correction not directly checkable from the paper alone.

Recommendation: Provide the explicit expression for $W_{\text{CIC}}(\mathbf{k})$ (discrete-grid form) used for deconvolution, and state any handling near the Nyquist frequency.

Key statements and references

- • **The simulations adopt the fiducial flat Λ CDM cosmology of the Quijote simulation suite, and the ensemble-averaged matter power spectrum $\langle P(k) \rangle$ from our GPU-accelerated PM runs is validated by direct comparison to the high-fidelity Quijote reference power spectrum used as an external benchmark.**
- *Reference(s):* Quijote
- • **Figure 1 compares the ensemble-averaged power spectrum from our simulations, $\langle P_{\text{sim}}(k) \rangle$, to the non-linear theoretical prediction from CAMB HaloFit, $P_{\text{CAMB,NL}}(k)$, showing that on large cosmological scales ($k < 0.03 h/\text{Mpc}$) the simulation reproduces the theoretical power spectrum to within approximately 5%, while underpredicting power by about 10–15% at intermediate scales ($0.03 \leq k \leq 0.1 h/\text{Mpc}$) and more strongly at smaller scales ($k > 0.1 h/\text{Mpc}$).**
- *Reference(s):* CAMB HaloFit
- • **These results confirm that our GPU-accelerated simulation provides reliable predictions on the large scales ($k < 0.1 h/\text{Mpc}$) most relevant for many cosmological analyses, such as studies of Baryon Acoustic Oscillations, while accurately capturing the expected limitations of the PM method at smaller scales when compared against the non-linear CAMB HaloFit theoretical model.**
- *Reference(s):* CAMB HaloFit

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper contains a small number of core formulas describing 2LPT initial conditions, a Fourier-space Poisson solve in a PM code, and a power-spectrum estimator with CIC and shot-noise corrections. There are no long derivations; the main analytic audit points are defini-

tion/normalization consistency (comoving vs physical variables, Fourier conventions, density vs density contrast) and symbol clarity.

Checked items

1. ✓ 2LPT displaced positions (Eq. (1), Sec. 2.2, p.3)

- **Claim:** Final particle positions are obtained by adding first- and second-order displacement fields to the initial grid (Lagrangian) position: $\vec{q}(\vec{x}) = \vec{x} + \vec{\Psi}^{(1)}(\vec{x}) + \vec{\Psi}^{(2)}(\vec{x})$.
- **Checks:** symbol/definition consistency, dimensional sanity
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** \vec{x} is the initial Lagrangian (grid) position as stated in Sec. 2.2, $\vec{\Psi}^{(1)}$, $\vec{\Psi}^{(2)}$ are displacement fields with dimensions of length
- **Notes:** Given the paper's own definitions, the mapping from initial grid position to displaced position is consistent and dimensionally sensible.

2. ✓ 2LPT peculiar velocity expression (Eq. (2), Sec. 2.2, p.3)

- **Claim:** Peculiar velocities are $\vec{v}(\vec{x}) = aH(a)f_1(a)\vec{\Psi}^{(1)}(\vec{x}) + aH(a)f_2(a)\vec{\Psi}^{(2)}(\vec{x})$.
- **Checks:** dimensional sanity, symbol/definition consistency
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** a is the scale factor (dimensionless), $H(a)$ has units 1/time, f_1, f_2 are growth-rate-like dimensionless factors as stated, $\vec{\Psi}$ has units of length
- **Notes:** Dimensionally consistent ($aH \times \text{length} \rightarrow \text{length}/\text{time}$). However, the paper does not define f_1, f_2 precisely, so only dimensional/symbolic consistency can be checked.

3. △ Fourier-space Poisson inversion (Eq. (3), Sec. 2.3, p.3)

- **Claim:** The gravitational potential is obtained in Fourier space by $\hat{\phi}(\vec{k}) = -4\pi G\hat{\rho}(\vec{k})/k^2$.
- **Checks:** algebra/sign check, definition consistency (source term and coordinates), sanity/edge cases ($k \rightarrow 0$)
- **Verdict:** UNCERTAIN; confidence: medium; impact: critical
- **Assumptions/inputs:** The real-space equation being inverted is implicitly $\nabla^2\phi = 4\pi G\rho$, Hats denote Fourier transforms, k^2 denotes $|\vec{k}|^2$
- **Notes:** Algebraic inversion/sign is consistent with $\nabla^2\phi = 4\pi G\rho$. But internal verifiability is blocked because the paper frames the simulation as cosmological/comoving (comoving volume; leapfrog with Hubble drag) yet does not specify whether ρ is physical density or density contrast, whether the mean

density is removed ($\mathbf{k} = \mathbf{0}$ handling), or whether factors are absorbed into ϕ . Without these definitions, the core force computation cannot be audited for internal consistency.

4. \triangle **Power spectrum estimator** (Eq. (4), Sec. 2.4, p.4)

- **Claim:** The matter power spectrum is estimated as $P(k) = \langle |\hat{\delta}(\vec{k})|^2 \rangle_k$ (bin-average over modes in a radial k -bin).
- **Checks:** definition/normalization consistency, dimensional sanity
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** $\delta(\vec{x})$ is a density contrast field assigned on a grid, $\hat{\delta}$ is obtained via a 3D FFT, $\langle \cdot \rangle_k$ denotes averaging over modes within a k -bin
- **Notes:** Structurally reasonable, but $P(k)$ depends on the FFT normalization convention (e.g., whether $\hat{\delta}$ includes $1/V$ factors). The paper does not define the Fourier convention, so the expression cannot be confirmed to produce a quantity with the intended units or to match the stated shot-noise subtraction form.

5. \checkmark **CIC deconvolution correction** (Sec. 2.4, p.4 (text following Eq. (4)))

- **Claim:** The measured power is corrected for CIC smoothing by dividing by the square of the CIC window function in Fourier space.
- **Checks:** symbolic method consistency
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** Same mass-assignment scheme (CIC) is used for field construction, Deconvolution uses $|W_{\text{CIC}}(\vec{k})|^2$
- **Notes:** The correction described is the standard symbolic form for undoing multiplicative window suppression in Fourier space. The exact window is not provided, so only the qualitative algebraic structure can be checked.

6. \triangle **Shot-noise subtraction term** (Sec. 2.4, p.4 (text following CIC correction))

- **Claim:** Shot noise is subtracted as $1/\bar{n}$, where \bar{n} is the mean particle number density.
- **Checks:** definition/normalization consistency, dimensional sanity
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** Particles sample a Poisson process around the underlying density field, δ and $\hat{\delta}$ normalization are compatible with a constant white-noise level $1/\bar{n}$
- **Notes:** Whether the correct constant is exactly $1/\bar{n}$ depends on how δ is defined and the Fourier/FFT normalization. The paper does not specify these, so this analytic correction cannot be verified internally.

7. \triangle **Ambiguity in particle/grid count notation** (Secs. 1, 2.1, 2.3, 3.1, 4 (multiple mentions))

- **Claim:** The simulation uses “ 512^3 particles” and a grid of “ 512^3 cells”.
- **Checks:** notation consistency
- **Verdict:** UNCERTAIN; confidence: high; impact: minor
- **Assumptions/inputs:** In 3D PM contexts, counts are often written as N^3 ; the text rendering may have dropped superscripts
- **Notes:** The document text is ambiguous between 512^3 and 5123. This is not an algebraic error per se, but it materially affects interpretability of subsequent analytic statements involving \bar{n} , grid Nyquist scale, and resolution-limited behavior.

Limitations

- The audit is limited to the provided PDF text/images; key implementation equations (e.g., equations of motion in comoving variables, leapfrog update formulas, and the exact Poisson source term) are not shown, preventing verification of the central dynamical system beyond the one-line Poisson inversion.
- Fourier-transform conventions (normalization, discrete k -grid definition) are not specified, limiting the ability to verify $P(k)$ and shot-noise expressions analytically.
- The CIC window function and any explicit deconvolution formula are not included, limiting checks to qualitative algebraic structure.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

10 candidate numeric checks were evaluated: 5 PASS and 5 UNCERTAIN, with 0 FAIL. Passes include ensemble-time multiplication ($20 \text{ s} \times 10 \approx 200 \text{ s}$), k -regime boundary consistency at 0.03 and 0.1 h/Mpc , ratio-to-percent consistency (0.96 corresponds to 4% deviation within a 5% target), and internal ordering/sanity checks for reported percentage ranges and cosmic-variance thresholds. Uncertainties mainly arise where the text provides ambiguous formatting (“5123” vs 512^3) or where only derived quantities can be computed without a stated reference value to compare against.

Checked items

1. Δ **C1_particles_count_cube** (Page 1 Abstract; Page 2 (end of intro); Page 6 Conclusions)
 - **Claim:** Simulation evolves 512^3 particles (written as 5123 in text).
 - **Checks:** power_of_integer
 - **Verdict:** UNCERTAIN
 - **Notes:** Computed $512^3 = 134,217,728$ particles, but the reported raw “5123” is ambiguous and no unambiguous explicit total particle count is provided to compare.

2. \triangle **C2_grid_cells_count_cube** (Page 2, Section 2.1 Simulation setup)
 - **Claim:** Force grid resolution is 512^3 cells (written as 5123 cells), matching particle number.
 - **Checks:** power_of_integer
 - **Verdict:** UNCERTAIN
 - **Notes:** Computed $512^3 = 134,217,728$ cells, but the reported raw "5123 cells" is ambiguous and no unambiguous explicit grid-cell total is provided to compare.
3. \checkmark **C3_ensemble_total_time** (Page 4, Section 3.1 Computational performance)
 - **Claim:** Each simulation completed in approximately 20 seconds; total time for 10-realization ensemble approximately 200 seconds.
 - **Checks:** parts_to_total
 - **Verdict:** PASS
 - **Notes:** $20\text{ s} \times 10 = 200\text{ s}$, matching the stated total within the stated approximate tolerance.
4. \triangle **C4_speedup_hours_to_seconds** (Page 4, Section 3.1 Computational performance; Page 6 Conclusions)
 - **Claim:** Wall-clock time reduced from approximately 5 hours on CPU to approximately 20 seconds on GPU.
 - **Checks:** unit_conversion_ratio
 - **Verdict:** UNCERTAIN
 - **Notes:** From the stated times, $5\text{ h} = 18,000\text{ s}$ and the implied speedup is $900\times$, but no explicit speedup factor is stated to verify.
5. \triangle **C5_time_step_size_in_scale_factor** (Page 3, Section 2.3 N-body evolution)
 - **Claim:** Evolved from $z = 127$ to $z = 0$ using 200 time steps uniform in scale factor a .
 - **Checks:** derived_step_size
 - **Verdict:** UNCERTAIN
 - **Notes:** Derived: $a_{\text{start}} = 1/(1 + 127) = 0.0078125$, $a_{\text{end}} = 1$, so $\Delta a = (1 - 0.0078125)/200 = 0.0049609375$. No stated Δa is provided to compare.
6. \triangle **C6_mean_number_density_from_particles_and_volume** (Page 1 Abstract; Page 2 Section 2.1; Page 4 Section 2.4 (shot noise $1/\bar{n}$))
 - **Claim:** Shot noise is $1/\bar{n}$; with 512^3 particles in a $(1000\text{ Mpc}/h)^3$ volume, \bar{n} can be computed.
 - **Checks:** derived_quantity
 - **Verdict:** UNCERTAIN

- **Notes:** Derived: $N = 512^3 = 134,217,728$; $V = 1000^3 = 1,000,000,000 \text{ (Mpc}/h)^3$; $\bar{n} = 0.134217728 \text{ (h/Mpc)}^3$; $1/\bar{n} = 7.450580596923827 \text{ (Mpc}/h)^3$. No explicit numeric \bar{n} or $1/\bar{n}$ is stated to verify.
7. ✓ **C7_k_regime_bounds_consistency** (Page 5 Section 3.2; Figure 1 caption; Page 6 discussion)
- **Claim:** Large-scale agreement claimed for $k < 0.03 \text{ h/Mpc}$; intermediate $0.03 \leq k \leq 0.1$; small scales $k > 0.1$; also claims reliable predictions on large scales ($k < 0.1 \text{ h/Mpc}$).
 - **Checks:** interval_consistency
 - **Verdict:** PASS
 - **Notes:** Boundaries are consistent: **0.03** matches the large/intermediate boundary and **0.1** matches the intermediate/small and 'reliable' threshold; inequality directions imply a contiguous partition with boundary handling as stated.
8. ✓ **C8_ratio_vs_percent_agreement** (Page 5 Section 3.2)
- **Claim:** Ratio on large scales is approximately **0.96**, achieving target of **5%** agreement.
 - **Checks:** percent_from_ratio
 - **Verdict:** PASS
 - **Notes:** Computed deviation from unity is $|1 - 0.96| = 0.04$ (**4%**), which is within a **5%** target. (The stored diff reflects $0.05 - 0.04 = 0.01$.)
9. ✓ **C9_underprediction_range_mid_scales** (Page 5 Section 3.2)
- **Claim:** At intermediate scales, systematic underprediction is on the order of **10–15%**.
 - **Checks:** range_sanity
 - **Verdict:** PASS
 - **Notes:** Range is internally consistent: $0.10 \leq 0.15$ and both values lie within $[0, 1]$.
10. ✓ **C10_cosmic_variance_vs_thresholds** (Page 6, Section 3.2 and Figure 2 caption)
- **Claim:** At largest scales ($k \leq 0.05 \text{ h/Mpc}$) standard deviation can exceed **20%** of mean; at smaller scales ($k \geq 0.1 \text{ h/Mpc}$) cosmic variance becomes negligible ($< 2%$).
 - **Checks:** inequality_threshold_consistency
 - **Verdict:** PASS
 - **Notes:** Ordering is logically consistent: $0.05 < 0.1$ and $0.20 > 0.02$.

Limitations

- Audit is based only on the provided parsed text; numeric values embedded in plots/images are not extracted or verified.
- Several key validation claims (power spectrum agreement/ratios) are qualitative or approximate without tabulated data, limiting fast recomputation checks.
- Ambiguity in typography (e.g., '5123' likely meaning 512^3) cannot be conclusively resolved without the original PDF formatting; checks treat it as a candidate for verification rather than a confirmed value.