



Full evidence edition / 28 September 2026

BGC87 / FOUR-SYSTEM COMPARISON

Scientific answers. Time to result.

A cross-field assessment of reference agreement, recorded delivery time and observed repeatability.

Linur Engine / Opus / Sol / Astra

Reference agreement

Delivery time

Observed repeatability

87

planned cases

13

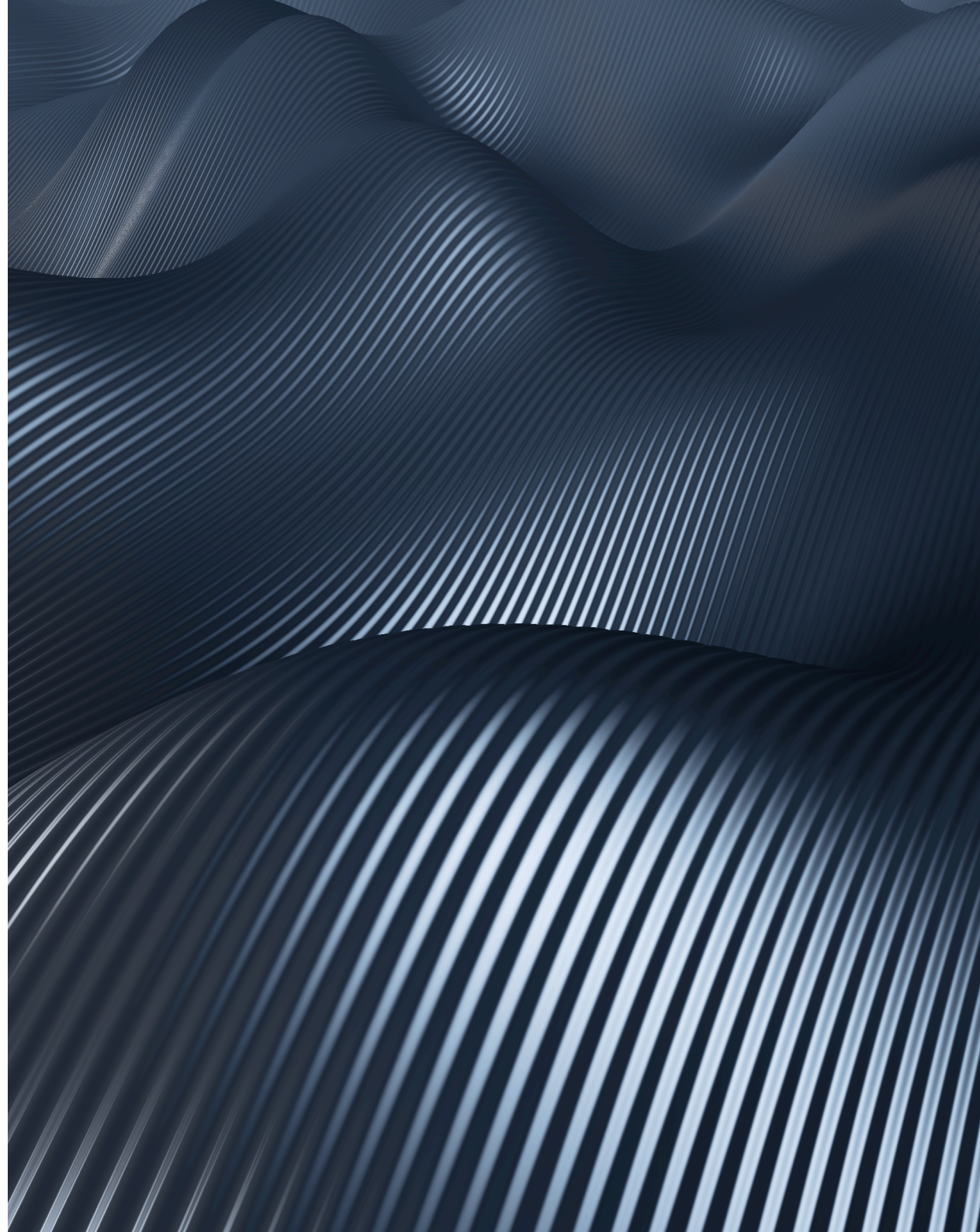
field groupings

4

systems

Deep-space alloy

Conceptual illustration, not measurement data or a model ranking.



Fields and comparison approach

Linur Engine, Opus, Sol and Astra. Reference agreement, delivery time and repeat observations remain separate. All cases retained.

Fields and tested bottlenecks

Calibration Numerical reference tasks 8 planned cases	Batteries Electrochemical storage 6 planned cases	2D materials Electronic materials 5 planned cases
Renewables Energy and catalysis 8 planned cases	Nuclear science Isotopes and nuclear energy 7 planned cases	Fusion Plasma-facing materials 6 planned cases
Quantum tech Quantum computing and sensing 3 planned cases	AI hardware AI and computing materials 11 planned cases	AI discovery Autonomous discovery 5 planned cases
Engineering Structural materials 5 planned cases	Biomedicine Biotechnology and medicine 11 planned cases	Chemistry Spectra and reaction networks 7 planned cases
Crystal discovery Crystal structure prediction 5 planned cases		

Why these questions matter

Materials properties, calibration, fits, spectra and trends connect calculations with devices and decisions.

The simulation challenge

Accuracy, model assumptions, convergence, computing time and resources all matter. No separate conventional-solver comparison was run.

INPUTS Sourced tasks Declared data and contracts	OUTPUTS Requested quantities One quantity is one requested output, not one task.
RULES Benchmark tolerance An output-specific allowed difference; not universal scientific truth.	EVIDENCE Relative difference Absolute error / absolute reference. Median within, then across cases.

[Global field charts](#) and task-specific evidence retain every produced result, regardless of protocol verdict. Reference differences, criterion margins and delivery time/repeatability are assessed separately. Field captions describe scope, not whole-field validation.

Performance by field / 1 of 7

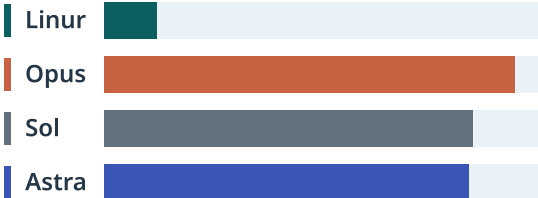
Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
 Calibration 8 planned tasks Quantum many-body reference; Main-group thermochemistry; Noncovalent interactions; Solid-state reproducibility; Lattice dynamics & thermal transport; Criticality & neutron transport; Liquid water structure & dynamics	Linur Opus Sol Astra 1 matched timed tasks.	0.326 s Linur: 4.30735e-4 relative-difference ratio. 202.02 s Opus: 0 relative-difference ratio. 225.47 s Sol: 0 relative-difference ratio. 196.73 s Astra: 0 relative-difference ratio. 4 cases / 51 matched scalar outputs. Case sources: [E01] [E02] [E03] [E04] [E05] [E06] [E07] [E08]
 Batteries 6 planned tasks Lithium-metal & solid-state batteries; Solid-state electrolytes; Cathode degradation & thermal runaway; Fast charging & lithium plating	Linur Opus Sol Astra 2 matched timed tasks.	0.306 s Linur: 0.0739431 relative-difference ratio. 360.25 s Opus: 0.0478905 relative-difference ratio. 324.85 s Sol: 0.0610685 relative-difference ratio. 160.48 s Astra: 0.0545425 relative-difference ratio. 6 cases / 218 matched scalar outputs. Case sources: [E09] [E10] [E11] [E12] [E13] [E14]

Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Performance by field / 2 of 7

Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
 2D materials 5 planned tasks Twistronics & quantum materials; 2D semiconductors for quantum technology; Graphene-based sensors	 5 matched timed tasks.	Linur: 0 relative-difference ratio. Opus: 0 relative-difference ratio. Sol: 0 relative-difference ratio. Astra: 0 relative-difference ratio. 5 cases / 43 matched scalar outputs. Case sources: [E15] [E16] [E17] [E18] [E19]
 Renewables 8 planned tasks Perovskite photovoltaics; Electrocatalysis; Thermoelectrics; Hydrogen storage	 8 matched timed tasks.	Linur: 9.1261e-7 relative-difference ratio. Opus: 9.1261e-7 relative-difference ratio. Sol: 9.1261e-7 relative-difference ratio. Astra: 9.1261e-7 relative-difference ratio. 8 cases / 65 matched scalar outputs. Case sources: [E20] [E21] [E22] [E23] [E24] [E25] [E26] [E27]

Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Performance by field / 3 of 7

Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
 Nuclear science 7 planned tasks Advanced reactor inventory & decay heat; Radiation damage in structural materials; Medical isotope production; Hydrogen isotope separation; Actinide chemistry & reprocessing	Linur Opus Sol Astra  5 matched timed tasks.	Linur: 8.57165e-16 relative-difference ratio. Opus: 5.63438e-16 relative-difference ratio. Sol: 7.1396e-16 relative-difference ratio. Astra: 7.1396e-16 relative-difference ratio. 5 cases / 36 matched scalar outputs. Case sources: [E28] [E29] [E30] [E31] [E32] [E33] [E34]
 Fusion 6 planned tasks Core plasma turbulence & confinement; Plasma-material interactions; Tritium breeding blanket; High-temperature superconducting magnets	Linur Opus Sol Astra  1 matched timed tasks.	Linur: 0.0278143 relative-difference ratio. Opus: 0.00762203 relative-difference ratio. Sol: 0.00602232 relative-difference ratio. Astra: 0.00878275 relative-difference ratio. 4 cases / 22 matched scalar outputs. Case sources: [E35] [E36] [E37] [E38] [E39] [E40]

Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Performance by field / 4 of 7

Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
 Quantum tech 3 planned tasks Superconducting qubit decoherence; Colour-centre spin qubits	Linur Opus Sol Astra 1 matched timed tasks.	0.215 s 262.82 s 295.79 s 354.80 s Linur: 3.04611e-15 relative-difference ratio. Opus: 2.07572e-16 relative-difference ratio. Sol: 1.79658e-16 relative-difference ratio. Astra: 1.79658e-16 relative-difference ratio. 3 cases / 27 matched scalar outputs. Case sources: [E41] [E42] [E43]
 AI hardware 11 planned tasks Post-silicon transistors; Non-volatile memory; Thermal management in 3D integrated circuits; Neuromorphic computing; EUV lithography & patterning; Wide-bandgap power electronics	Linur Opus Sol Astra 6 matched timed tasks.	0.167 s 637.19 s 201.08 s 294.91 s Linur: 0.0064094 relative-difference ratio. Opus: 0.00521392 relative-difference ratio. Sol: 0.0064094 relative-difference ratio. Astra: 0.0064094 relative-difference ratio. 10 cases / 92 matched scalar outputs. Case sources: [E44] [E45] [E46] [E47] [E48] [E49] [E50] [E51] [E52] [E53] [E54]

Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Performance by field / 5 of 7

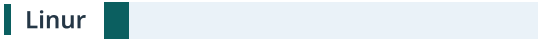
Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
 AI discovery 5 planned tasks AI-driven materials discovery; AI-driven protein design; Autonomous closed-loop experimentation	No shared timed cases Recorded intervals / planned cases: Linur 0 / 5 Opus 4 / 5 Sol 4 / 5 Astra 4 / 5 Timing details in the Annex.	No shared four-system scalar summary Produced / requested outputs: Linur 2 / 11 Opus 10 / 11 Sol 10 / 11 Astra 10 / 11 Details in the Annex: case-specific assessments, not missing calculations. Case sources: [E55] [E56] [E57] [E58] [E59]
 Engineering 5 planned tasks High-entropy alloys; Additive manufacturing; Corrosion & passivation; Rare-earth-lean permanent magnets	Linur Opus Sol Astra 2 matched timed tasks.	0.160 s 298.91 s 160.02 s 139.12 s Linur: 0.0453782 relative-difference ratio. Opus: 0.0490232 relative-difference ratio. Sol: 0.06258 relative-difference ratio. Astra: 0.0576795 relative-difference ratio. 4 cases / 65 matched scalar outputs. Case sources: [E60] [E61] [E62] [E63] [E64]

Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Performance by field / 6 of 7

Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
 Biomedicine 11 planned tasks Structure-based drug design; Drug resistance; Enzyme design; Antimicrobial resistance; Biopharmaceutical formulation; Nanomedicine & drug delivery	No shared timed cases Recorded intervals / planned cases: Linur 0 / 11 Opus 11 / 11 Sol 11 / 11 Astra 9 / 11 Timing details in the Annex.	No shared four-system scalar summary Produced / requested outputs: Linur 8 / 83 Opus 82 / 83 Sol 83 / 83 Astra 64 / 83 Details in the Annex: case-specific assessments, not missing calculations. Case sources: [E65] [E66] [E67] [E68] [E69] [E70] [E71] [E72] [E73] [E74] [E75]
 Chemistry 7 planned tasks Homogeneous catalysis; Combustion & sustainable fuels; Excited-state dynamics; Spectroscopic prediction	Linur  0.166 s Opus  282.63 s Sol  159.16 s Astra  220.06 s 2 matched timed tasks.	Linur: 0.0375277 relative-difference ratio. Opus: 0.00300449 relative-difference ratio. Sol: 0.00148485 relative-difference ratio. Astra: 0.00783832 relative-difference ratio. 6 cases / 23 matched scalar outputs. Case sources: [E76] [E77] [E78] [E79] [E80] [E81] [E82]

Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Performance by field / 7 of 7

Reference agreement, recorded time and missing evidence remain distinct.

Field and tested scope	Recorded delivery time Less is better / log axis, 0.1-1000 seconds	Reference evidence Less is better / relative difference, not accuracy %.
Crystal discovery 5 planned tasks High-temperature superconductors; Gas separation & carbon capture; Metastable phase discovery	Linur Opus Sol Astra 2 matched timed tasks.	0.168 s Linur: 0 relative-difference ratio. 339.19 s Opus: 6.82753e-10 relative-difference ratio. 178.59 s Sol: 8.7814e-16 relative-difference ratio. 116.49 s Astra: 0 relative-difference ratio. 2 cases / 26 matched scalar outputs. Case sources: [E83] [E84] [E85] [E86] [E87]

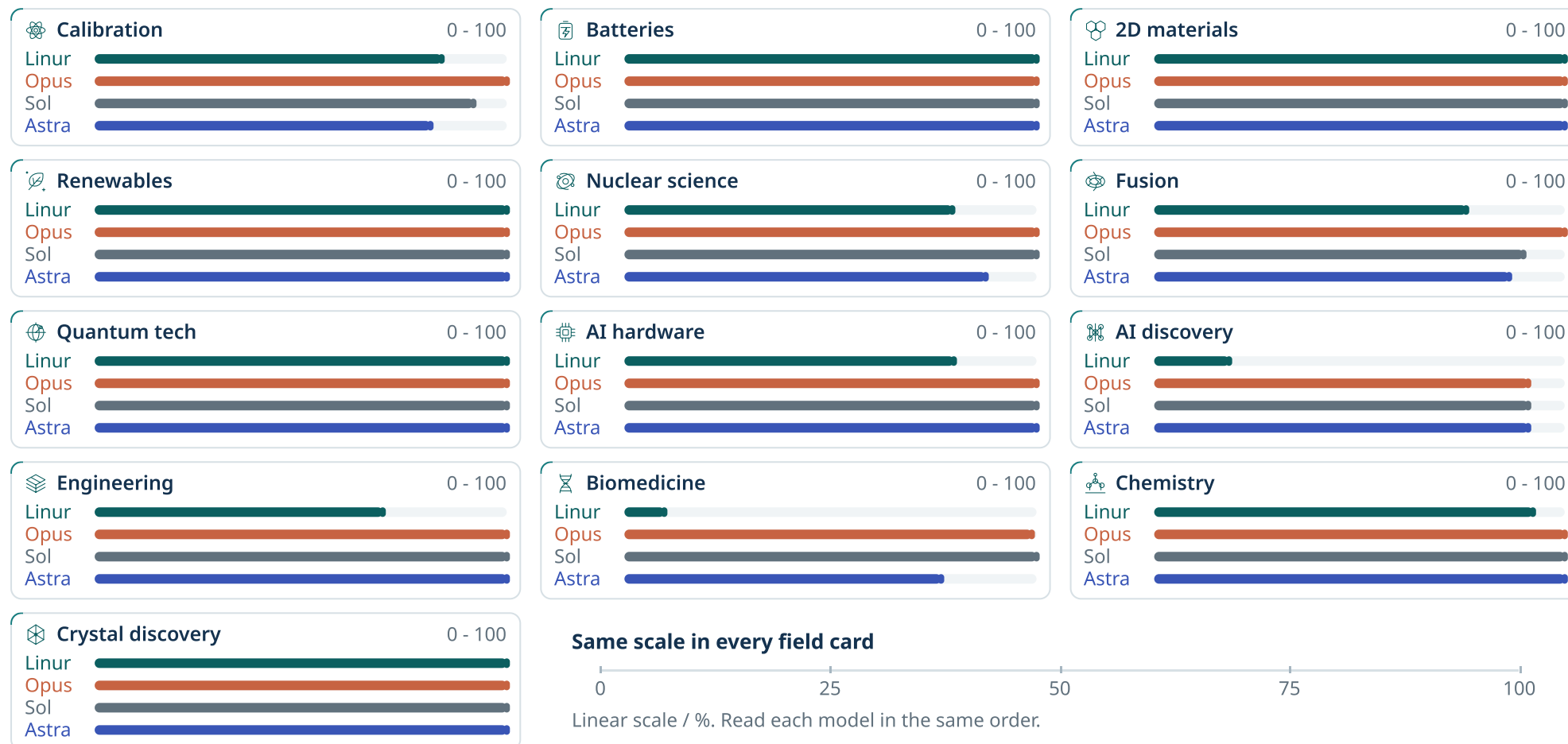
Scalar summaries are medians within, then across cases. Missing a common cohort does not mean no results were assessed. Timing boundaries differ; recovery never inherits the original unsuccessful interval.

Received outputs by field

Produced / requested primary outputs; coverage is not correctness.

Linur Opus Sol Astra Linear scale / %

More is better / for output coverage, not accuracy



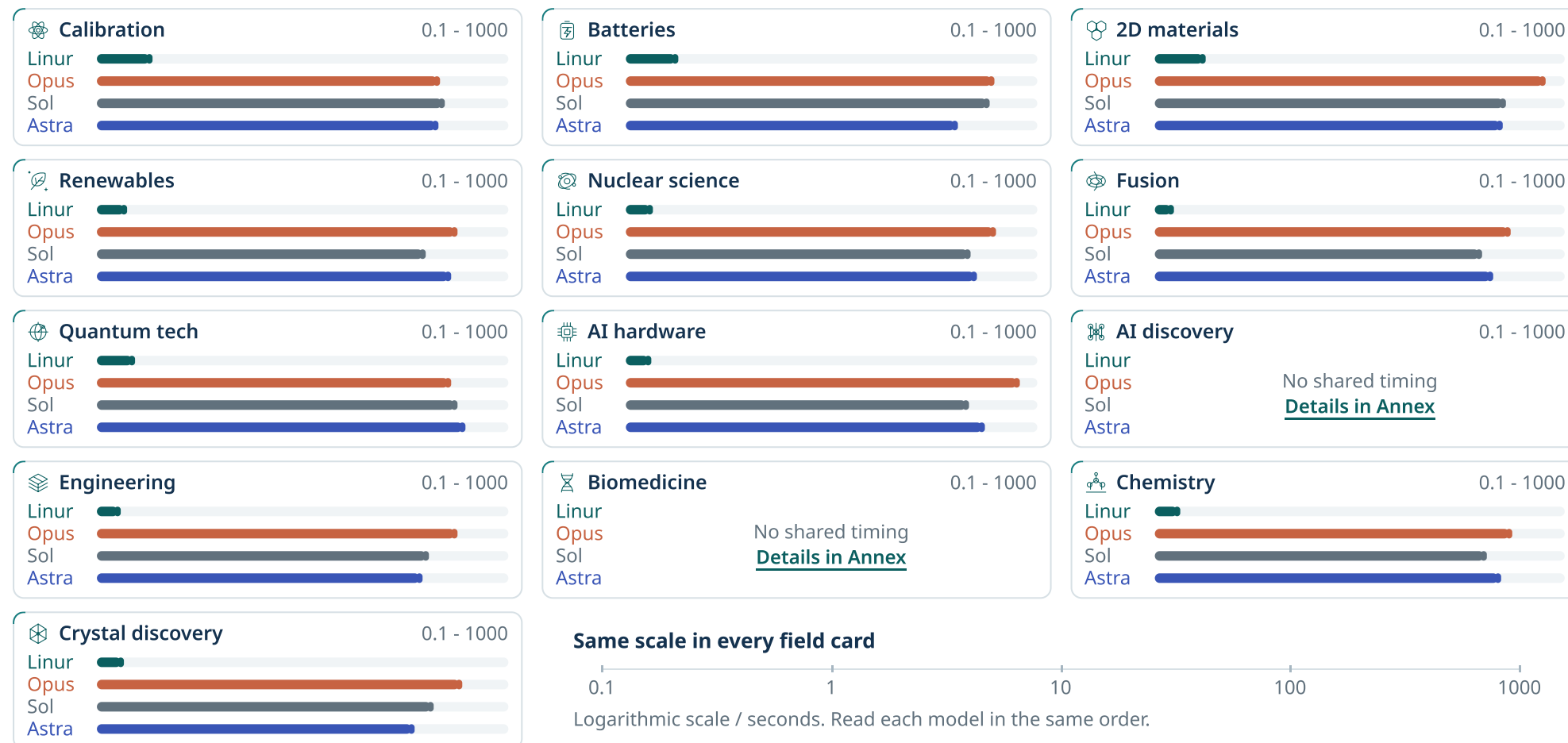
Bars show produced / requested primary outputs as a percentage. An output can contain many matrix elements; coverage is not scientific correctness. [Exact values, cohorts and definitions: Annex.](#)

Time to answer by field

Median seconds on the same complete-output cases for all four systems.

■ Linur ■ Opus ■ Sol ■ Astra Log scale / seconds

Less is better / for recorded time to answer



Median seconds on shared complete-output cases. Recorded timing boundaries differ: not equal-hardware compute time. Missing Linur intervals are never inferred. [Exact values, cohorts and definitions: Annex.](#)

Scientific correctness evidence by field

Scalar reference difference; lower is closer to the declared reference.

Linur Opus Sol Astra Log scale / ratio

Less is better / for closeness to the declared reference



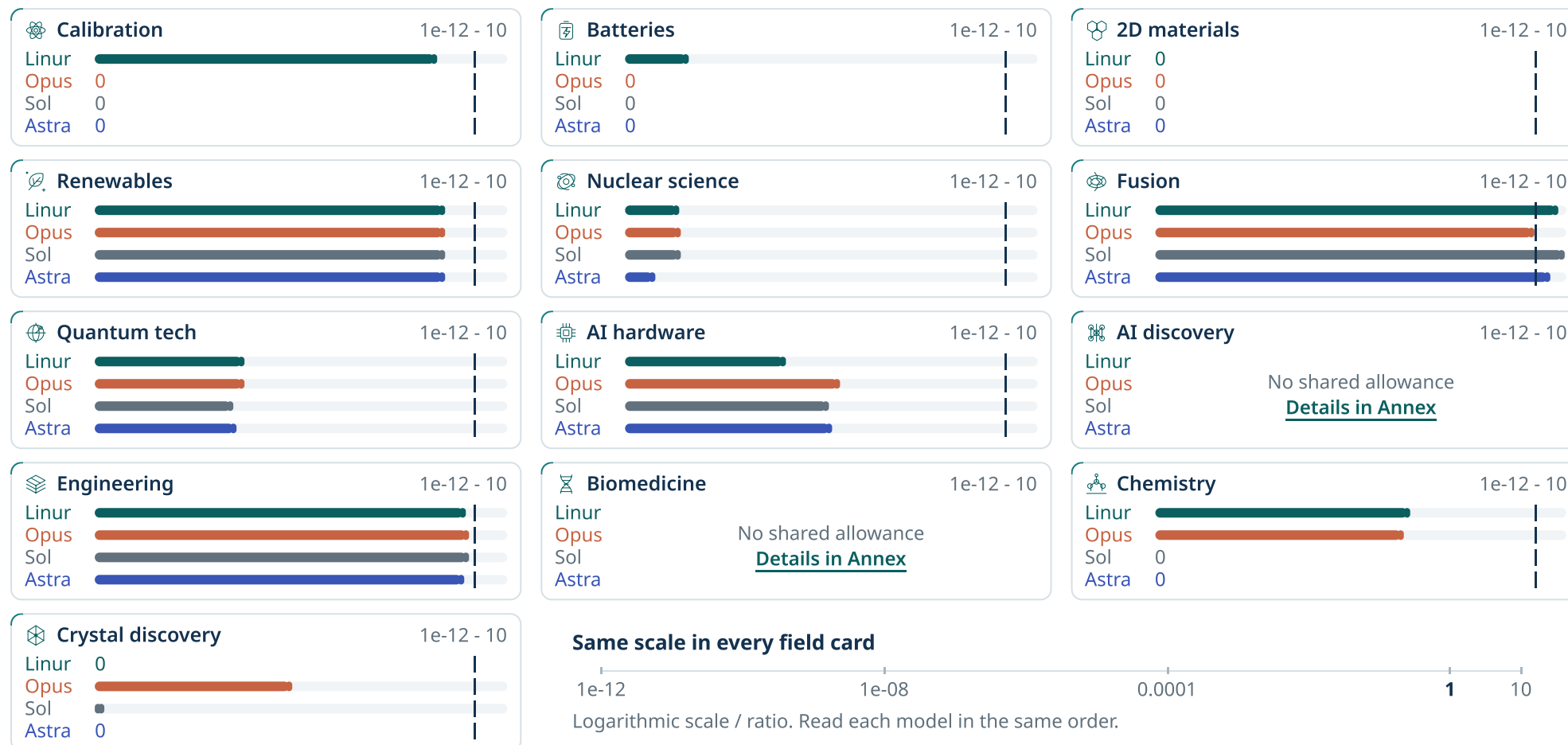
Absolute error / absolute reference; median within each case, then across cases. Not accuracy percentages. Medians can hide large errors; censored and undefined ratios stay separate. Exact zero is written as 0 instead of a log-scaled bar. [Exact values, cohorts and definitions: Annex.](#)

Distance from the original allowances

Error / allowance; below 1 is inside the numerical allowance.

■ Linur ■ Opus ■ Sol ■ Astra Log scale / ratio

Less is better / for error relative to its allowance



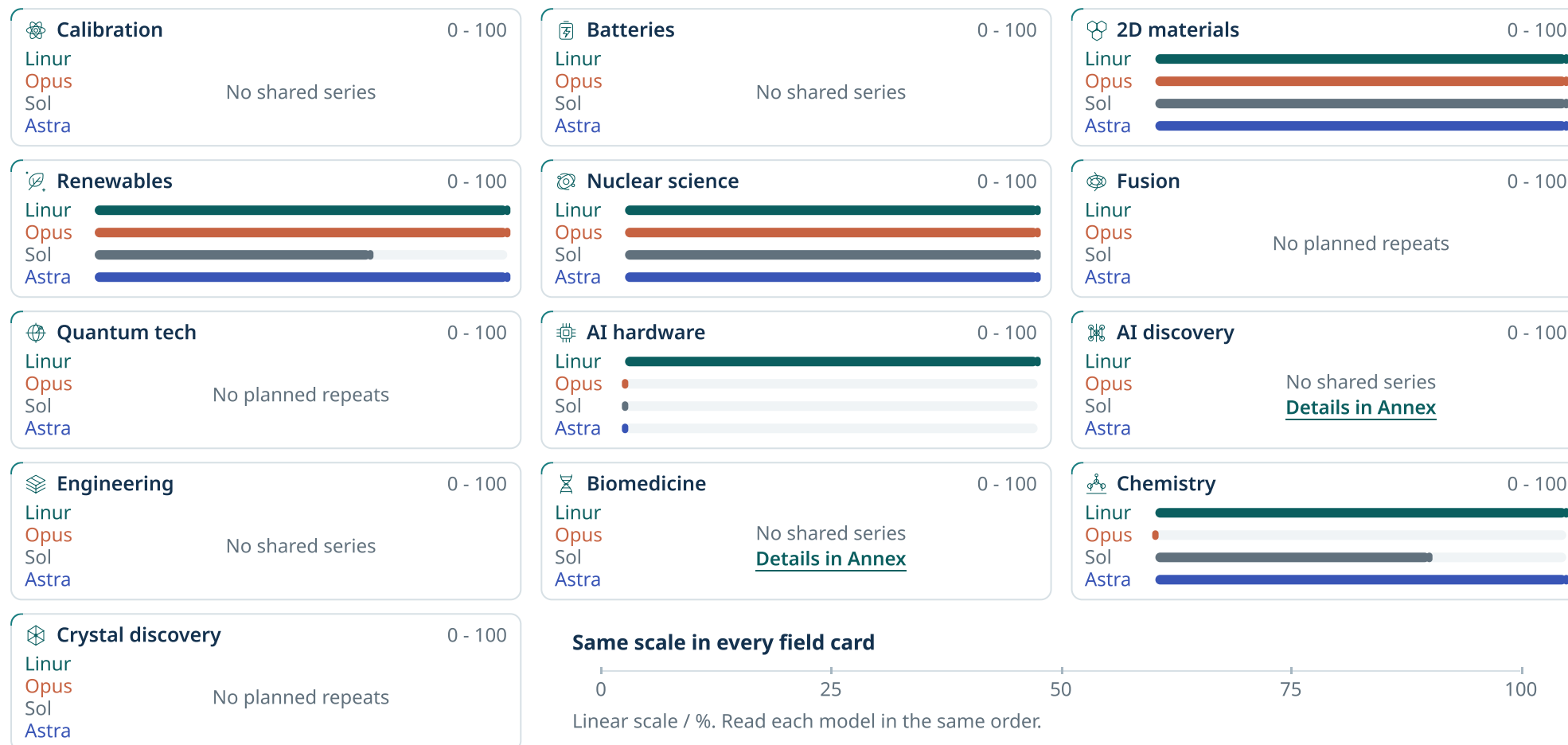
Median error / allowance within each case, then across cases. The bold line at 1 marks the original allowance, not a new acceptance rule. Exact zero is written as 0 instead of a log-scaled bar. [Exact values, cohorts and definitions: Annex.](#)

Observed repeat agreement by field

Exact agreement across three recorded trials on the same shared output series.

Linur Opus Sol Astra Linear scale / %

More is better / for observed consistency, not correctness



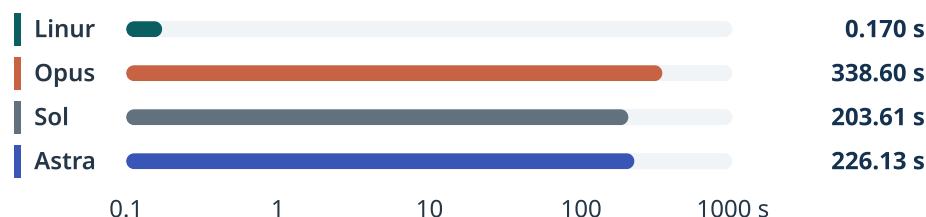
Percentage of shared complete three-trial series with exact agreement. No shared series and no planned repeats are different gaps; agreement is not correctness or controlled reproducibility. [Exact values, cohorts and definitions: Annex.](#)

Result assessment and delivery time

FAIL and schema mismatch do not exclude a produced numerical observation.

🕒 Time to answer

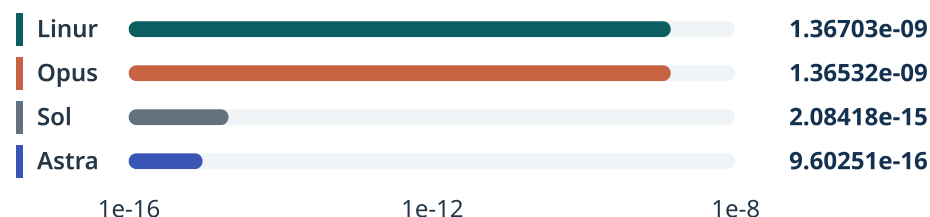
Less is better. 35 matched timed tasks; median seconds.



Different boundaries; not common-hardware compute time.

📊 Reference difference

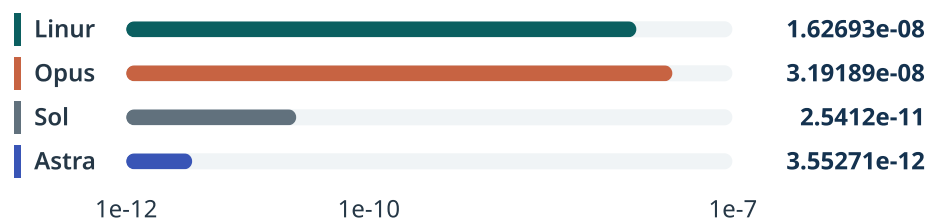
Less is better. 57 cases / 668 shared scalar outputs.



Median of case-median relative errors. Log ratio axis, not accuracy %.

📏 Distance from allowance

Less is better. 42 cases / 366 shared tolerance outputs.



Median of case-median error/allowance ratios. Limit 1 lies beyond this log axis.

🔄 Observed repeat agreement

More is better. 34 shared complete three-value series.



Exact observed agreement; not correctness or controlled reproducibility.

Each panel has its own population and scale. Categories, vectors and eight censored conductivity bounds use their own metrics in the annex. Medians can hide large individual errors; no universal score or composite system ranking is created.

What the results establish

Assess the observations

4,724 produced result cells across primaries and repeats. 4,715 support a reference-based numerical or categorical assessment; nine have no supplied reference.

Preserve protocol decisions

PASS / FAIL remains the historical protocol label. Signed errors and distance from the original allowance are shown even for FAIL. A negative margin is inside the allowance; positive is outside.

Resolve the format barrier

All 121 recovered conductivity values are assessed numerically: 113 exact references and eight censored bounds. Their frozen exact-text score remains unresolved, not rewritten.

What correctness means here

Closer agreement with a specified reference is measurable evidence, not independently established scientific truth. Reference uncertainty, source resolution, physical interpretation and model limitations still matter. No threshold is invented for an originally descriptive rule.

Speed and repeatability remain separate

Timing uses its recorded boundaries, not a PASS filter. 780 of 1,012 planned output-repeat series have at least two produced observations; partial series and runtime identity limits remain explicit.

No missing output, time, resource measurement or repeat has been synthesized. This is not a common-hardware efficiency or universal superiority claim.

Provenance and limits

Unchanged source evidence

Original and recovery packages, all 492 trial records, and historical scores remain intact. Internal recovery is first-attempt evidence by user instruction, not witnessed execution attestation.

Declared versus derived

Source-declared log, vector, selection and ranking metrics are identified. Additional signed differences, criterion margins and descriptive summaries are labelled analyst calculations.

Outside the denominator

Nine saved B-minus-A values are assessed against explicit derived A/B reference contrasts. They do not increase the 881 requested outputs per system.

Scientific qualifications stay visible

Upper-bound consistency is not exact-reference agreement. Numeric text is parsed only for descriptive arithmetic; no response is converted or categorical rule changed. Sentinel mismatches, ordering, status provenance and unresolved cutoff ties remain explicit.

The S032 temperature interpretation/equality discrepancy and source-resolution caveats remain. Original Linur accepted-attempt IDs stay null. The historical 45-case / 420-check protocol cohort is preserved as historical evidence, not used as the new assessment filter.

Limits of the comparison

Thirty-five cases have common complete-output timing. Recovery component intervals are reported separately, without replacing missing end-to-end time. Agreement and recorded repeat stability do not establish architecture, mechanism, universal capability or independently validated correctness.

FULL EVIDENCE EDITION. Reference-sensitive material. Not independent scientific peer review.