

The Joules-per-Solution Metric: Definition, Measurement Protocol, and Anti-Gaming Provisions for Honest Computational Benchmarking

The Joules-per-Solution Metric: Definition, Measurement Protocol, and Anti-Gaming Provisions for Honest Computational Benchmarking

Author: QNFO | **Date:** 2026-07-27 | **Status:** Draft — Phase 5 **Series:** Joules-per-Compute Universal Benchmark (JPCUB) — Paper P0 **License:** QNFO Unified License Agreement (QNFO-ULA)

Abstract

The global computational infrastructure spans seven distinct paradigms — from classical CPUs to quantum processors, from AI accelerators to neuromorphic chips — yet no universal metric exists for comparing their computational value. Every paradigm benchmarks itself on its own terms: quantum volume for quantum computing, FLOPs per second for AI accelerators, SPEC scores for CPUs, and Power Usage Effectiveness for data centers. None of these are honest proxies for what physics charges: energy. This paper proposes [my conjecture] the **joules-per-solution metric** — the total system-level energy required to produce a correct solution to a computational task at a specified correctness threshold — as the universal, physics-grounded, falsifiable arbiter of computational advantage. We provide a formal definition with closed-form measurement equations, a measurement protocol covering computation, cooling, power conversion, networking, and amortized manufacturing energy, and a comprehensive anti-gaming framework with adversarial validation, pre-registration requirements, and Pareto frontier reporting. We survey 14 existing domain-specific benchmarks — NeuroBench, ML.ENERGY, SPECpower, Green500, MLPerf Power, ULPMark, TPC-Energy, Green Graph 500, CoreMark-Pro, CloudSuite, HPCG, and the Sustainable Quantum Computing carbon framework — and demonstrate that while each measures energy within its silo, none provides cross-domain comparability. We argue that the absence of a universal, physics-grounded benchmark has enabled systematic overstatement of computational value, most visibly in quantum computing (\$35B invested, zero commercially viable machines meeting the joules-per-solution criterion), and propose an 8-item calibration register with 2028–2035 checkpoints to enable independent verification of our framework's predictive power.

Keywords: computational benchmarking, energy efficiency, joules per operation, Landauer limit, honest computation, quantum computing, neuromorphic computing, AI energy, green computing

1. Introduction

1.1 The Fifteen Orders of Magnitude

The computational infrastructure of the twenty-first century spans an extraordinary range of energy efficiencies. At one extreme, the Landauer limit — $kT \ln 2 \approx 2.9 \times 10^{-21}$ J at room temperature — defines the irreducible thermodynamic cost of erasing one bit of information [Landauer1961]. At the other extreme, fault-tolerant gate-model quantum computers, burdened by cryogenic cooling at millikelvin temperatures, control electronics operating at gigahertz frequencies, and error-correction overheads of 10^2 to 10^6 physical qubits per logical qubit, consume approximately 10^{12} to 10^{15} times the Landauer limit per logical operation.

Between these extremes lie six additional computational paradigms, each operating within its own energy-efficiency envelope:

- **Classical CMOS CPUs** operate at roughly 10^3 times the Landauer limit, constrained by Dennard scaling breakdown and the end of Moore's Law.
- **AI accelerators (GPUs, TPUs)** consume 10^4 to 10^5 times Landauer per operation, with frontier model training runs exceeding 50 GWh — enough to power 5,000 homes for a year.
- **Data centers**, after two decades of Power Usage Effectiveness (PUE) optimization from 2.0 to 1.1, face diminishing returns on facility efficiency while per-computation costs remain opaque.
- **Neuromorphic computing** claims a 100–1,000× energy advantage over digital CMOS for specific task classes, operating at potentially 10 – 10^2 times Landauer — three to five orders of magnitude more efficient than GPUs — but lacks standardized benchmarking to substantiate these claims.
- **Thermodynamic and reversible computing** approaches the Landauer limit theoretically but has produced no commercially viable near-Landauer prototype at scale.
- **Edge and IoT devices** operate under battery constraints where joules-per-inference directly determines device lifetime, yet no cross-platform benchmark exists for comparing inference energy across heterogeneous hardware.

The critical observation is not only that these paradigms span fifteen orders of magnitude in energy efficiency. It is that **investment is inversely correlated with energy efficiency**. The least energy-efficient paradigm — gate-model quantum computing — has absorbed approximately \$35 billion in global investment [paper-the-qubit-delusion], while the most energy-efficient paradigms — neuromorphic and thermodynamic computing — receive orders of magnitude less funding.

1.2 The Blind Spot

This inversion exists because every computational paradigm benchmarks itself on its own terms. Quantum computing uses "quantum volume," a metric defined by IBM that increases with the number of qubits — regardless of whether those qubits can perform useful computation [2403.02240]. AI accelerators use FLOPs per second, a metric that ignores the exponential growth in training energy documented for frontier models [2109.05472]. Data centers report PUE, a facility-level metric that reveals nothing about the computational value delivered per joule. Classical CPUs benchmark on SPEC scores and Geekbench, which measure wall-clock time and ignore energy consumption entirely.

None of these metrics are **honest proxies for what physics charges**. The universe does not care about FLOPs. It does not care about quantum volume. It does not care about PUE. It cares about joules. Every gate transition, every memory access, every cooling cycle, every photon detection event costs energy. A computational paradigm that claims "advantage" while ignoring its joule cost is claiming to violate the laws of thermodynamics [paper-physics-of-computation].

This paper proposes the corrective: **the joules-per-solution metric**.

1.3 Core Claim

Joules-per-solution is the only honest arbiter of computational advantage. For any computational task T with a correctness threshold ε , the computational advantage of system A over system B is defined as:

$$\text{Advantage}_T(A, B) = \frac{E_B(T, \varepsilon)}{E_A(T, \varepsilon)}$$

where $E_X(T, \varepsilon)$ is the total system-level energy consumed by system X to produce a solution to T within correctness bound ε . System A has advantage over B if and only if $\text{Advantage}_T(A, B) > 1$.

This claim is falsifiable. It would be disconfirmed if, for a representative sample of ≥ 100 real-world computational tasks spanning ≥ 4 of the 7 paradigm domains, a different metric (wall-clock time, FLOPs, or quantum volume) predicted procurement outcomes more accurately than joules-per-solution, where procurement outcome is defined as the system chosen by an independent evaluator given total cost of ownership data including energy costs.

2. Background and Prior Work

2.1 Physical Foundations

The physics of computation establishes three mathematically derived limits relevant to this work.

The Landauer bound [established] (1961): any logically irreversible operation — such as erasing a bit — dissipates at minimum $kT \ln 2$ joules of energy as heat [Landauer1961]. This is not an engineering limit but a consequence of the Second Law of Thermodynamics. Reversible computing can, in principle, approach zero energy dissipation per operation by avoiding logical irreversibility, but all current practical computing systems — classical and quantum — operate far above this bound [1806.10183].

The Margolus–Levitin theorem [established] (1998): a quantum system with average energy E requires at least $\Delta t \geq h/(4E)$ to evolve between two distinguishable states [MargolusLevitin1998]. This bounds the maximum computational speed of any quantum device and, combined with the Landauer bound, establishes a minimum energy-time product for computation.

The Bremermann limit [established] (1962): a material system of mass m cannot process information faster than $mc^2/h \approx 1.36 \times 10^{50}$ bits per second per kilogram [Bremermann1962]. This is the ultimate speed limit, far beyond current technology, but it establishes the envelope within which all computation must occur.

These three bounds — thermodynamic (Landauer), dynamical (Margolus–Levitin), and throughput (Bremermann) — define the honest boundaries within which any claim of computational advantage must be assessed [paper-physics-of-computation].

2.2 Domain-Specific Benchmarking [established]: A Survey of Fourteen Competitors

We surveyed fourteen existing benchmarks that measure energy or energy-proxy metrics for specific computational domains. Table 1 summarizes our findings.

Table 1: Fourteen Domain-Specific Energy Benchmarks

Benchmark	Domain	Year	Metric	Cross-Domain?	Anti-Gaming?	Living?
NeuroBench [2304.04640]	Neuromorphic	2023	Accuracy + throughput + power	No	No	No
ML.ENERGY [2505.06371]	ML Inference	2025	Inference energy	No	No	No
SPECpower_ss2008	Server CPUs	2007	Throughput/watt (Java)	No	No	No
Green500	HPC	2007	FLOPs/watt (LINPACK)	No	No	No

Benchmark	Domain	Year	Metric	Cross-Domain?	Anti-Gaming?	Living?
MLPerf Power	ML Training/Inference	2019	Power during benchmark	No	No	No
EEMBC ULPMark	Microcontrollers	2014	Energy per benchmark	No	No	No
TPC-Energy	Database	2010	Watts/transaction	No	No	Deprecated
Green Graph 500	Graph Processing	2010	Energy per graph traversal	No	No	No
EEMBC CoreMark-Pro	Embedded/IoT	2015	Energy per benchmark	No	No	No
CloudSuite	Cloud Computing	2011	Energy per workload	No	No	No
HPCG	HPC	2014	FLOPs/watt (CG solver)	No	No	No
SQC Carbon [2408.05679]	Quantum Computing	2024	Carbon footprint (lifecycle)	No	No	No
Approximate Computing Survey [2307.11128]	Cross-cutting	2023	Energy-quality tradeoffs	Partially	No	No
Google ML Carbon	AI/ML	2022	Training carbon (white paper)	No	No	No

No existing benchmark provides cross-domain comparability. Every benchmark is designed for a specific paradigm, workload, or device class. None use "joules per solution" as a universal metric. None embed anti-gaming provisions. None propose a living benchmark protocol with annual updates and community contribution.

2.3 The Joules-per-Solution Gap

Despite the existence of fourteen domain-specific energy benchmarks, no framework exists to answer the following questions:

1. Is a Google TPU v5 more energy-efficient than an Intel Xeon for training a 7B-parameter language model, measured in joules per converged model?
2. Is an IBM quantum processor more energy-efficient than a classical Monte Carlo simulation for computing molecular ground-state energies to within 1.6 mHa chemical accuracy?
3. Is a Loihi 2 neuromorphic chip more energy-efficient than an NVIDIA H100 GPU for real-time video object detection at 95% accuracy?
4. How does the energy efficiency of lattice-based post-quantum cryptography compare to classical RSA-2048 on the same embedded hardware?

These are not hypothetical questions. They are the questions that a procurement officer, a research funding agency, or a climate-conscious engineer should be asking before committing resources. The absence of a common yardstick means these questions go unasked — and the resources go to the paradigm with the best marketing, not the best physics.

3. The Joules-per-Solution Metric [my conjecture]

3.1 Formal Definition

Definition 1 (Task). A computational task $T = (I, O, \delta)$ consists of an input space I , an output space O , and a correctness function $\delta : O \times O \rightarrow \{0, 1\}$ that determines whether a computed output $\hat{o}$ is correct relative to a reference output o^* .

Definition 2 (Solution at Correctness Threshold ε). A computational system S produces a solution to task T at correctness threshold ε if, when presented with an input $i \in I$ sampled from a representative input distribution $\mathcal{D}_T$, system S produces output $\hat{o}$ such that:

$$\Pr_{i \sim \mathcal{D}_T} [\delta(\hat{o}, o^*(i)) = 1] \geq \varepsilon$$

where $o^*(i)$ is the reference correct output for input i .

Definition 3 (Total System Energy). The total system energy $E_S(T, \varepsilon)$ consumed by system S to produce a solution to task T at correctness threshold ε is:

$$E_S(T, \varepsilon) = \sum_{k=1}^K E_k(T, \varepsilon)$$

where K energy components are measured:

Component	Symbol	Description
Computation	E_{comp}	Energy consumed by processing elements during task execution
Memory	E_{mem}	Energy consumed by memory hierarchy (registers, cache, DRAM, storage)
I/O and Networking	E_{io}	Energy consumed by data movement between components and across networks
Cooling	E_{cool}	Energy consumed by cooling infrastructure to maintain operating temperature
Power Conversion	E_{conv}	Energy lost in AC-DC and DC-DC power conversion
Amortized Manufacturing	E_{mfg}	Energy cost of manufacturing amortized over expected device lifetime operations

Definition 4 (Joules-per-Solution). The joules-per-solution for system S on task T at correctness threshold ε is:

$$J_S(T, \varepsilon) = E_S(T, \varepsilon)$$

3.2 Measurement Protocol

The measurement protocol is designed to be reproducible, auditable, and resistant to gaming. It consists of five phases:

Phase 1: Task Specification. The task T is defined with:

- Input distribution $\mathcal{D}_T$ and representative sample size $N \geq 100$
- Correctness function δ with automated verification
- Pre-registered correctness threshold(s) $\varepsilon \in \{0.80, 0.85, 0.90, 0.95, 0.99\}$

Phase 2: System Instrumentation. Energy measurement at the wall-plug or DC rail level using calibrated power meters with $\pm 1\%$ accuracy or better. For components where direct measurement is infeasible (e.g., amortized manufacturing energy), physics-based models using published lifecycle assessment data provide lower and upper

bounds.

Phase 3: Warm-Up and Steady-State. System is run at representative load for ≥ 10 minutes before measurement begins to reach thermal steady-state. Power measurements are sampled at ≥ 1 Hz.

Phase 4: Task Execution and Measurement. The task is executed on N inputs. For each input i , power $P_i(t)$ is recorded from task start to completion. Energy per input is:

$$E_i = \int_{t_{\text{start}}}^{t_{\text{end}}} P_i(t) dt$$

Phase 5: Reporting. Results are reported as a Pareto frontier of $(E_S(T, \epsilon), \epsilon)$ across the five correctness thresholds. No single number is reported — the full frontier prevents cherry-picking a favorable threshold.

3.3 Total System Energy Accounting

The total system energy accounting is the key differentiator between JPCUB and existing benchmarks. Most existing benchmarks measure chip power (TDP) or facility power (PUE). JPCUB measures total system energy including:

Computation (E_{comp}): The energy consumed by processing elements — CPU cores, GPU SMs, quantum gates, neuromorphic neurons — during task execution. Measured at DC rail level for digital systems; includes idle power during task execution (not just active power).

Memory (E_{mem}): Energy consumed by all levels of the memory hierarchy — registers, L1/L2/L3 cache, DRAM, NVMe storage, HBM. For data-intensive tasks (AI training, database queries), memory energy can exceed computation energy by 2–5 $\times$.

I/O and Networking (E_{io}): Energy consumed by data movement. For distributed computing (data centers, HPC clusters), network switch and router energy is included proportionally. For cryptographic systems, the energy of transmitting ciphertexts is included — a critical factor for post-quantum cryptography where ciphertexts are 10–100 $\times$ larger than classical equivalents [1903.04570].

Cooling (E_{cool}): Energy consumed by active cooling. For air-cooled data centers ($\text{PUE} \approx 1.1\text{--}1.3$), cooling adds 10–30% overhead. For quantum computers operating at millikelvin temperatures, dilution refrigerator power (≈ 15 kW for operational QC systems) is included — this multiplies the chip-level energy by factors of 10–100 $\times$ [2408.05679].

Power Conversion (E_{conv}): Energy lost in AC-DC rectification (typically 5–10% loss) and DC-DC voltage regulation (typically 10–15% loss at the board level). Total power conversion losses typically add 15–25% to DC-level measurements.

Amortized Manufacturing (E_{mfg}): The energy cost of manufacturing the hardware, amortized over expected lifetime operations. For a server CPU with a 5-year lifetime, manufacturing energy adds approximately 10–20% to operational energy. For quantum processors requiring isotopic purification of silicon-28 or fabrication of superconducting circuits at nanoscale precision, manufacturing energy may be significantly higher [2408.05679].

3.4 Reporting Standards

All joules-per-solution measurements must report:

1. **Measurement date** — hardware and software configurations evolve; measurements are valid as of the date reported

2. **System configuration** — processor model, memory configuration, cooling method, software versions
 3. **Measurement instrument** — power meter model, accuracy class, calibration date
 4. **Task specification** — input distribution, sample size, correctness function, reference implementation
 5. **Raw data** — time-series power data for all N inputs, made publicly available
 6. **Pareto frontier** — $J_S(T, \varepsilon)$ for all five ε thresholds
 7. **Energy component breakdown** — $E_{\text{comp}}, E_{\text{mem}}, E_{\text{io}}, E_{\text{cool}}, E_{\text{conv}}, E_{\text{mfg}}$ individually reported
 8. **Uncertainty bounds** — 95% confidence intervals for each measurement
-

4. Anti-Gaming Provisions [speculative]

The history of computational benchmarking is a history of gaming. Every benchmark that becomes consequential — SPEC, LINPACK, MLPerf — has been optimized against in ways that inflate reported performance relative to real-world utility. A joules-per-solution benchmark that lacks anti-gaming provisions would be particularly vulnerable: energy measurements can be manipulated by adjusting clock frequencies, disabling components, cherry-picking inputs, or reporting chip power instead of system power.

4.1 Pre-Registration Requirement

All joules-per-solution measurements must be pre-registered before execution. The pre-registration includes: the task specification $(I, O, \delta, \mathcal{D}_T)$, the correctness threshold(s) ε , the sample size N , the measurement methodology, and the system configuration.

Pre-registration prevents post-hoc threshold tuning — the practice of measuring first, then selecting the ε that produces the most favorable result. Once pre-registered, the measurement must be executed as specified and results must be reported for ALL pre-registered thresholds, even if unfavorable.

4.2 Adversarial Validation

Every joules-per-solution claim must be independently reproducible. The measurement protocol requires: open-source task implementations, publicly available input datasets, calibrated measurement instruments with documented accuracy, and auditable data collection scripts.

The burden of proof is on the claimant, not the auditor. If a claim cannot be independently reproduced, it is not a claim — it is a press release.

4.3 Pareto Frontier Reporting

The most common form of benchmark gaming is selecting the operating point that optimizes the reported metric while ignoring other dimensions. Reporting joules-per-solution at $\varepsilon = 0.99$ might favor one system (high accuracy, high energy); reporting at $\varepsilon = 0.80$ might favor another (low accuracy, low energy). The Pareto frontier requirement eliminates this degree of freedom: results must be reported for all five thresholds, and the full frontier must be presented.

4.4 Energy Component Audit

Each of the six energy components ($E_{\text{comp}}, E_{\text{mem}}, E_{\text{io}}, E_{\text{cool}}, E_{\text{conv}}, E_{\text{mfg}}$) must be individually reported with measurement or modeling methodology. An auditor can challenge any component: "Your cooling model assumes a coefficient of performance of 3.0, but the manufacturer's datasheet specifies 2.5." The component-level breakdown enables targeted challenges rather than blanket dismissal.

4.5 Living Benchmark Protocol

Benchmarks become stale. SPEC2006 is not comparable to SPEC2017. A joules-per-solution measurement from 2026 is not directly comparable to one from 2030 — hardware, software, and measurement instruments evolve. The living benchmark protocol addresses this with:

- **Annual measurement updates** — systems must be re-measured annually with current hardware/software configurations
- **Versioned methodology** — the measurement protocol is versioned (JPCUB v1.0, v2.0, ...) with documented changes
- **Community contribution** — hardware vendors, researchers, and independent auditors can submit measurements following the protocol
- **Deprecation schedule** — measurements older than 2 years are flagged as potentially stale; measurements older than 5 years are deprecated

5. Cross-Domain Comparison Framework [speculative]

5.1 The Comparative Atlas

The joules-per-solution metric enables, for the first time, a direct energy-efficiency comparison across all seven computational paradigms. Figure 1 illustrates the conceptual framework: each paradigm occupies a region in the 2D space of (correctness threshold ϵ , joules per solution J), and systems within each paradigm form a frontier of achievable (ϵ, J) pairs.

The Comparative Atlas (to be fully populated in JPCUB Paper P9) places all seven paradigms on a common scale:

Paradigm	Approximate J Range (relative to Landauer)	Dominant Cost Driver
Thermodynamic (theoretical)	$1\text{--}10^2 \times \text{Landauer}$	Adiabatic switching overhead
Neuromorphic	$10\text{--}10^2 \times \text{Landauer}$	Leakage current, routing
CMOS CPUs	$10^3\text{--}10^4 \times \text{Landauer}$	Dynamic switching, leakage
AI Accelerators	$10^4\text{--}10^5 \times \text{Landauer}$	Memory bandwidth, scaling
Data Centers	$10^5 \times \text{Landauer}$	Cooling, networking, utilization
Post-Quantum Cryptography	$10^4\text{--}10^6 \times \text{Landauer}$	Key/signature size, bandwidth
Fault-Tolerant QC	$10^{12}\text{--}10^{15} \times \text{Landauer}$	Cryogenic cooling, QEC overhead

5.2 The Investment Inversion

The most striking feature of this atlas is that **investment is inversely correlated with energy efficiency**. Quantum computing, at $10^{12}\text{--}10^{15}$ times the Landauer limit, has received approximately *35 billion*. Neuromorphic computing, at $10\text{--}10^2$ times Landauer, has received perhaps 1–2 billion. Thermodynamic computing, closest to the theoretical optimum, has received maybe \$100 million.

This is not a coincidence. It is a consequence of the absence of a universal benchmark. Without joules-per-solution, investment flows to the paradigm with the most compelling narrative, not the best physics. The JPCUB framework is designed to correct this market failure.

5.3 Cross-Domain Consilience

The joules-per-solution metric exhibits a structural isomorphism across disciplines [artifacts/consilience-gate]. In physics, it is thermodynamic efficiency ($\eta = W/Q$). In computer science, it is algorithmic $\times$ hardware cost. In biology, it is metabolic cost per adaptive behavior — the human brain achieves 10^{16} synaptic operations per second on 20 W, approximately 2×10^{-15} J per "operation," three orders of magnitude more efficient than the best CMOS [2604.04727]. In sociology, it is resource cost per institutional decision — the structural analogue of financial auditing standards (GAAP/IFRS) applied to computational claims.

The invariant across all six domains is: **self-reported metrics without independent, physics-grounded verification create systematic overestimation of value**. Joules-per-solution is the metric that physics enforces regardless of what humans choose to measure.

6. Calibration Register [speculative]

To enable independent verification of our framework's predictive power, we pre-register the following predictions. Each prediction has a defined checkpoint date and falsification condition.

ID	Prediction	Checkpoint	Status
CAL-01	No gate-model quantum computer will solve a commercially relevant problem at lower joules-per-solution than the best classical alternative	2030-12-31	[PENDING]
CAL-02	At least one major cloud provider (AWS/Azure/GCP) will publish joules-per-inference data for ML services	2028-12-31	[PENDING]
CAL-03	Neuromorphic computing will achieve $\geq 50\times$ CMOS energy advantage for ≥ 1 benchmark task	2030-12-31	[PENDING]
CAL-04	PQC migration will increase TLS handshake energy by $\geq 5\times$ for at least one measured deployment	2028-12-31	[PENDING]
CAL-05	This paper (JPCUB P0) will be cited by ≥ 10 external papers by 2030	2030-12-31	[PENDING]
CAL-06	At least one EU member state will include compute energy reporting in procurement regulations	2030-12-31	[PENDING]
CAL-07	Reversible computing will not achieve near-Landauer operation ($> 0.5kT \ln 2$) at any commercially meaningful scale	2035-12-31	[PENDING]
CAL-08	ARM/RISC-V server market share will exceed 30% by 2030	2030-12-31	[PENDING]

7. Limitations and Known Objections

7.1 The Correctness Threshold Problem

The strongest objection to the joules-per-solution metric is that the correctness threshold ε is doing too much work. Two systems may produce solutions at different accuracy levels: system A achieves 95% accuracy at 100 J per inference; system B achieves 85% accuracy at 1 J per inference. Which is "better"?

Our response is threefold: (a) the Pareto frontier reporting requirement eliminates this ambiguity — both ($J = 100, \varepsilon = 0.95$) and ($J = 1, \varepsilon = 0.85$) are reported, and the user chooses based on their accuracy requirement; (b) the pre-registration requirement prevents post-hoc threshold selection to favor a particular system;

(c) for tasks where a "correct" answer is unambiguous (e.g., factoring a 2048-bit integer), ϵ is binary and the problem disappears.

7.2 The Measurement Feasibility Problem

Measuring the total system energy of a quantum computer — including cryogenic cooling, control electronics at room temperature, and amortized manufacturing energy — requires access to hardware that is proprietary and often classified. How can JPCUB measure what it cannot access?

We do not claim to directly measure proprietary hardware. Our approach is: (a) provide physics-based lower bounds (Landauer limit, Margolus–Levitin speed limit) that no system can beat; (b) provide TDP-based upper bounds from published specifications; (c) invite hardware vendors to submit BETTER numbers if they believe our bounds are unfair. This is the "auditor's burden" principle: the burden of proof is on the claimant, not the auditor. If IBM believes its quantum computer is more energy-efficient than our Landauer-based lower bound suggests, IBM can publish a joules-per-solution measurement following this protocol.

7.3 The Adoption Chicken-and-Egg Problem

A universal benchmark without adoption is a paperweight. Adoption without a benchmark is impossible. This chicken-and-egg problem has killed previous cross-domain benchmarking efforts (e.g., TPC-Energy).

Our response: (a) JPCUB does not require universal adoption to be useful — even partial adoption by a single major cloud provider or procurement agency creates a competitive dynamic where others must follow; (b) the Green500 succeeded because HPC energy costs became the dominant constraint — we argue that the same dynamic is emerging across ALL computational paradigms as energy costs rise, climate regulation tightens, and the quantum computing bubble faces scrutiny; (c) the living benchmark protocol ensures that JPCUB remains relevant as hardware evolves, avoiding the SPEC2006-to-SPEC2017 stagnation problem.

7.4 The Quantum Computing Objection

Quantum computing advocates may object that joules-per-solution is an unfair metric because quantum computers are designed for problems where classical computers are exponentially slower, and the energy comparison should therefore account for the exponential speedup.

This objection misunderstands the joules-per-solution metric. If a quantum computer genuinely provides an exponential speedup for a commercially relevant problem, it WILL win on joules-per-solution — because the classical alternative will consume exponentially more energy. The metric is not anti-quantum; it is pro-honest-computation. If quantum wins, JPCUB reports that honestly. If quantum loses — as all currently published evidence suggests [[paper-the-qubit-delusion](#), [paper-manifesto-honest-computation](#)] — JPCUB reports that honestly too.

8. Conclusion

The global computational infrastructure is the largest energy-consuming technology system in human history, yet it operates without a universal, physics-grounded benchmark for comparing the value it delivers. Every paradigm — quantum, AI, classical, neuromorphic, cryptographic — measures itself on its own terms, and every paradigm's self-assessment overstates its value.

The joules-per-solution metric is the corrective. It is falsifiable. It is physics-grounded. It includes total system energy — computation, memory, I/O, cooling, power conversion, and amortized manufacturing. It embeds anti-gaming provisions — pre-registration, adversarial validation, Pareto frontier reporting, component-level audit, and a

living benchmark protocol.

The publication of this metric is not an academic exercise. It is an Archimedean lever. As demonstrated in the accompanying Deep Research Cascade (JPCUB Phase 4), the absence of a cross-domain energy benchmark is the binding constraint on honest computational investment. Removing that constraint — by publishing, open-sourcing, and demonstrating the joules-per-solution framework — changes the probability landscape for paradigm shifts worth hundreds of billions of dollars in reallocated R&D.

The universe does not care about FLOPs. It cares about joules. It is time our benchmarks did too.

References

See `artifacts/jpcub-bibliography.bib` for the complete 140-entry bibliography supporting this paper. Key citations reproduced below for reader convenience:

- Landauer 1961: Irreversibility and heat generation in the computing process
 - Margolus–Levitin 1998: The maximum speed of dynamical evolution
 - Bremermann 1962: Optimization through evolution and recombination
 - NeuroBench (2304.04640): Framework for benchmarking neuromorphic computing
 - ML.ENERGY (2505.06371): Toward automated inference energy measurement
 - SQC (2408.05679): Sustainable quantum computing — carbon benchmarking
 - Approximate Computing Survey (2307.11128): Energy-quality tradeoffs
 - Generalized Reversible Computing (1806.10183)
 - Fault-tolerant QC with constant overhead (2512.02760)
 - Thermodynamic recycling in QC (2601.07522)
 - Compute and energy trends in DL inference (2109.05472)
 - Neuromorphic computing for low-power AI (2604.04727)
 - The Qubit Delusion, The Physics of Computation, The Problem-Substrate Mapping, Manifesto for Honest Computation (QNFO, 2026)
-

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest: The authors declare no competing financial interests. The QNFO advocates for honest computational benchmarking as a public good. This paper is published under the QNFO Unified License Agreement to ensure open access.

Data Availability: All search queries, literature data, classification results, and BibTeX entries are available in the JPCUB GitHub repository at github.com/rwnq8/joules-per-compute-benchmark and archived on Cloudflare R2 at qnfo-releases/releases/2026/07/joules-per-compute-benchmark/.

Use of Artificial Intelligence: This paper was written with AI assistance for literature search, data processing, and initial drafting. All claims were verified against cited sources by human review. The AI system operated under the QNFO Research Integrity Mandate — all output is factual and evidence-grounded.