Energy and Carbon Footprint of the x3d2 CFD Solver on AI-First and HPC-Class GPUs

Irufan Ahmed
Department of Aeronautics
Imperial College London
London, United Kingdom
irufan.ahmed04@imperial.ac.uk

***Abstract*—Estimates of the energy and carbon cost of computation are almost never checked against measurement. We instrument an incompressible Computational Fluid Dynamics (CFD) solver across three GPUs and a server CPU. While the AI-first accelerator uses 1.27–1.40× lower energy than either HPC-class GPU, hardware efficiency is second-order: shifting a job's start time across 12 months alters emissions by 28×. Four standard carbon-accounting tools, all correctly configured, agree within 1.3× on that accelerator while sitting 1.9–2.9× above measured wall-socket energy—demonstrating that agreement does not imply accuracy. Reported carbon is dictated by measurement boundaries and scheduling, not hardware.**

***Index Terms*—CFD, carbon footprint, HPC, GPU**

I. Introduction

x3d2 [1] is a high-order, compact finite-difference Navier–Stokes solver rebuilt for heterogeneous hardware. Its spatially implicit schemes require ∼150 tridiagonal solves per timestep. A distributed algorithm exploiting diagonal dominance avoids MPI all-to-all communication, leaving only the pressure Poisson FFT. Its CUDA and OpenMP backends enable direct comparison.

II. Methodology

We evaluated the Taylor–Green Vortex ($Re = 1600$, $t = 20$, zero I/O) across three resolutions, repeating each configuration three to ten times. Hardware comprised an AMD EPYC 7443 host (CPU baseline) with two A100s, two identical NVIDIA DGX Sparks (GB10: Blackwell GPU), and an NVIDIA GH200 (Grace + H100) evaluated alongside the 124-run core campaign to test whether findings generalise.

GB10 delivers higher FP32 throughput than A100 but lower memory bandwidth; cuBLAS GEMM executes FP64 52× slower than FP32 with just two FP64 cores per multiprocessor; A100 and GH200 match both rates via FP64 tensor cores.

Cross-platform metrics report energy per working device: the idle second A100 draws 33.5 W. Power was sampled at 5 Hz via NVML, supplemented by RAPL and BMC telemetry on the x86 host. GB10 lacks both, so a plug meter measures its wall socket.

III. Results

A. GPU Offload and Node Placement

Offloading from 24 EPYC cores to one A100 reduced wall time 20-fold and whole-node energy 12-fold. During CPU-only runs, two idle A100s consumed 28% of node energy. Boundary selection governs multi-GPU metrics: adding a second A100 yields a 1.10× speedup, but increases working-device energy by 14% while saving 5% whole-node: the second card was already idling, so a shorter run amortises fixed power.

B. Platform Comparison and FP64

The A100 finishes 5.5× faster, yet GB10 uses 1.40× lower energy at the working device. Despite a 47× FP64 throughput deficit against A100, the solver's performance penalty tracks the 5.98× memory-bandwidth ratio rather than the arithmetic shortfall. The GH200 reinforces this finding: with 2.6× the A100's bandwidth, it completes 1.80× faster at near-identical energy per cell-step, indicating extra bandwidth buys time, not joules.

C. Single-Precision and Unit Variability

Single precision cuts GPU energy by 24–27% at 128^3 across all three GPUs, as halving operand width directly halves memory traffic. At 256^3, savings diverge (44% on GB10, 40% on GH200, 36% on A100) independently of FP64 capability, which spans 157× between them. Savings exceed speedups by 8–20 points: single precision cuts power as well as runtime, so time-proportional models under-report them. Enstrophy stays within 0.01% of double precision. Two identical GB10 units differ by 7.4% in GPU energy (58.75 vs 63.08 Wh, $n = 5$ vs 8) for a 0.07% wall-time difference; the wall socket agrees (1.067×). An attached monitor had raised one unit's energy by 24% with no runtime change.

D. Carbon-Aware Scheduling and Tool Accuracy

Shifting a GB10 job across 12 months of UK grid carbon intensity alters emissions by 28× (2.4 to 69.3 gCO_2e)—a swing 20× larger than the GB10-versus-A100 energy gap. Scheduling dominates hardware selection.

Even with verified SLURM GPU accounting, standard tools fail: GreenAlgorithms4HPC (static carbon intensity) and the API-driven GUILT, ARCHER2 jobemissions and GRACE-HPC [2]–[5] overestimate GB10 wall-socket energy by 1.9–2.9×, yet span 0.64–1.11× of measured whole-node energy on the A100 node. Consensus does not imply accuracy: half the tools lack GPU telemetry, relying on static CPU TDP or

whole-node constants, and none can resolve the 7.4% variance between identical units.

IV. Conclusions

Bandwidth-bound CFD on AI-first architectures carries far less penalty than FP64 benchmarks suggest, and precision savings hold across architectures. Environmental tracking remains broken: tools compound hardware-blind constants with static carbon averages. Accounting abstractions must shift to per-machine calibration and probabilistic distributions.

Acknowledgement

Supported by the Department of Aeronautics, Imperial College London. Claude (Anthropic, Opus 5) assisted with drafting; the author edited and verified the text. All measurements, analysis and conclusions are the author's own.

References

[1] xcompact3d contributors, "x3d2: next-generation high-fidelity CFD solver." [Online]. Available: https://github.com/xcompact3d/x3d2 (accessed Aug. 7, 2026).

[2] G. Roe, K. Jonathan, and the SCD Energy Efficiency Team, "GUILT: Green Usage Impact Logging Tool." [Online]. Available: https://github.com/SCD-Energy-Efficiency-Team/guilt (accessed Aug. 7, 2026).

[3] Cambridge Sustainable Computing Lab, "GA4HPC: Green Algorithms for high performance computing." [Online]. Available: https://github.com/Cambridge-Sustainable-Computing-Lab/GreenAlgorithms4HPC (accessed Aug. 7, 2026).

[4] ARCHER2, "emissions-tools: emissions estimates on ARCHER2." [Online]. Available: https://github.com/ARCHER2-HPC/emissions-tools (accessed Aug. 7, 2026).

[5] E. Ayliffe, "GRACE-HPC: Green Resource for Assessing Carbon & Energy in HPC." [Online]. Available: https://github.com/Elliot-Ayliffe/GRACE-HPC (accessed Aug. 7, 2026).