COIA-Ω Photonic AI Accelerator

Manufacturing-Ready 3× Energy-Efficient AI Co-Processor

De-risked design for next-generation data centres

15 August 2025

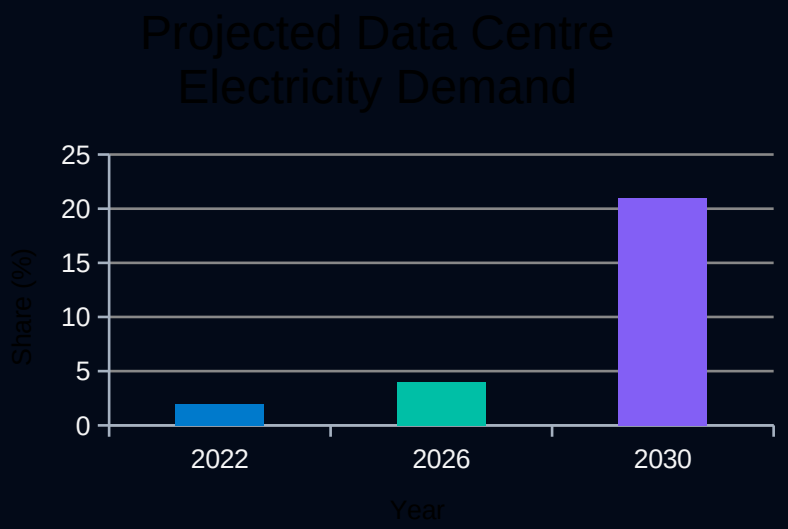
Executive Summary

- COIA-Ω is a photonic co-processor engineered to overcome the AI energy wall, performing matrix-vector operations directly with light.
- Designed as a PCIe Gen5 accelerator, it targets ≥3× energy-per-operation improvement versus NVIDIA’s H100 for key workloads.
- Implementation-ready: includes a 12-layer PCB stack, firmware boot sequences and a complete register map.

[1] [2]

The AI Energy Wall

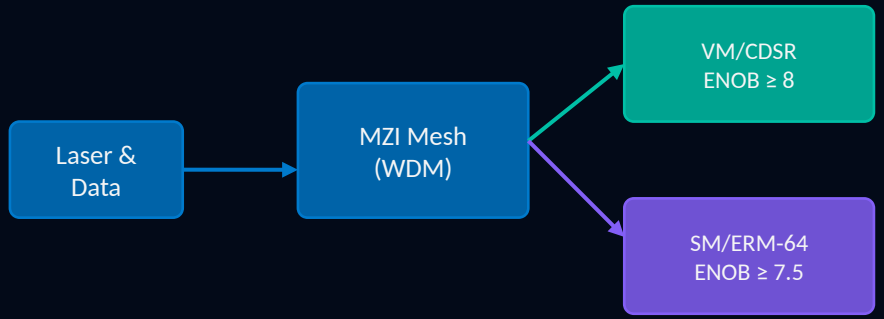
- AI model complexity and token throughput grow exponentially, driving unprecedented electricity demand.
- Data-centre electricity demand could double between 2022 and 2026 with AI adoption.
- By 2030, data centres may consume ≈21% of global electricity, eclipsing many industries.



[1] [2]

COIA-Ω Architecture

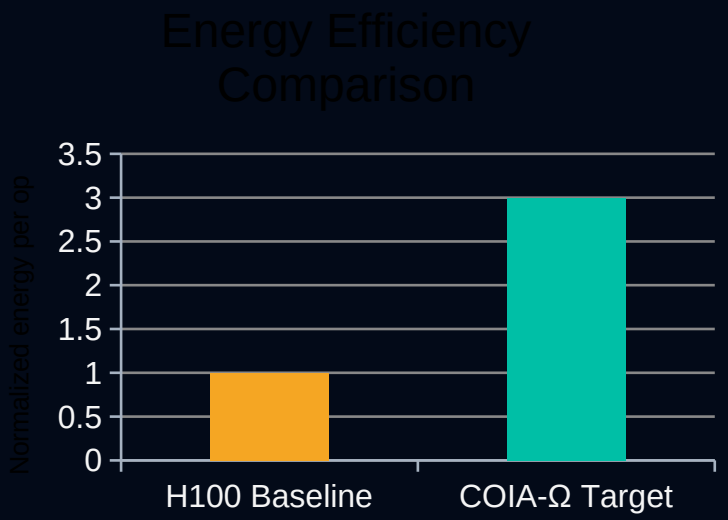
- Thin-film $LiNbO_3$ on SiN Mach-Zehnder mesh performs matrix-vector operations with coherent light.
- 8×100 GHz WDM channels enable massive parallelism across wavelengths.
- Dual readout modes: VM/CDSR (integrate-and-dump, ENOB ≥ 8) and SM/ERM-64 (continuous streaming, ENOB ≥ 7.5).
- Predictive thermal and phase control holds stability at ≤2° RMS and eliminates silent errors.



[1] [2]

Performance & Reliability

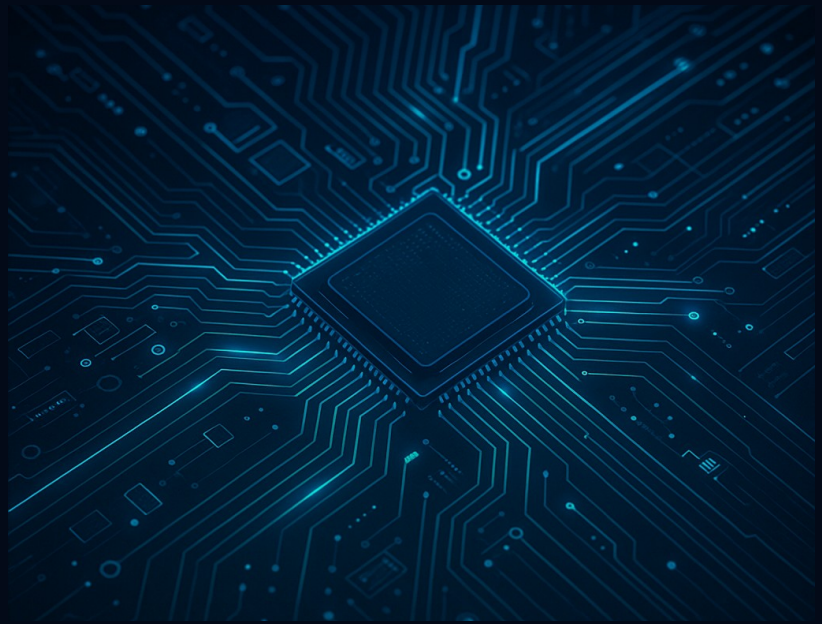
Metric	Target
Energy efficiency	≥3× vs H100
Insertion loss	≤5 dB (P95)
MTBF	≥1×10^{7} hours
Silent errors	$<10^{-15}$ per decision
Scalability	400G Ethernet & CXL 3.0



[1] [2]

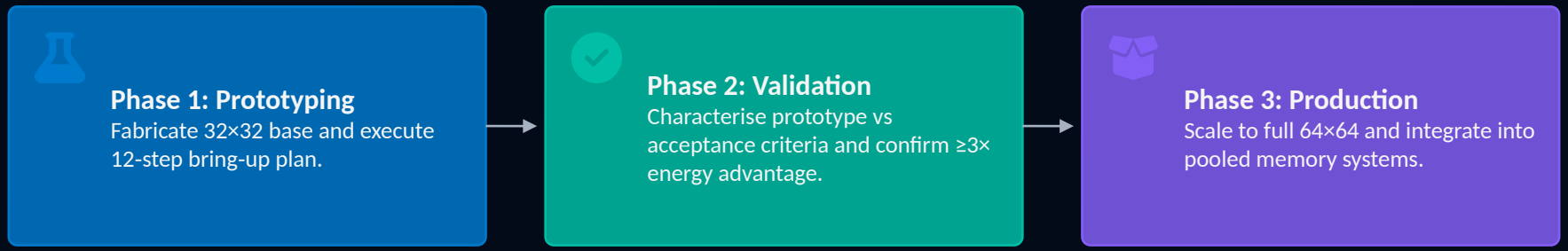
Implementation Maturity

- Form factor: PCIe full-length/full-height (312×112 mm); 12-layer PCB; board input 12 V DC.
- Power delivery: multiple rails (R0: +12 V, R1: +5 V_LAS, R2: +3.3 V, R3: +3.3 V_D, etc.) with controlled sequencing and decoupling.
- Register map: fully enumerated addresses with per-output CDSR_CFG, timing and status registers for deterministic control.



[1] [2] [3]

Deployment Roadmap



[1]

Conclusion

- COIA-Ω demonstrates that photonic computing can shatter the AI power wall by delivering ≥3× energy efficiency without sacrificing precision.
- Implementation readiness is evidenced by a complete low-level binder, from PCB and power to firmware and register maps.
- With a clear roadmap, this de-risked architecture is poised to redefine performance-per-watt in next-generation data centres.

[1] [2]