OPTICAL INTERCONNECTS

As a Critical Enabler for Next-Generation AI Infrastructure

By James Altucher and Michael Murray

July 2026

ABSTRACT

The scaling of artificial intelligence infrastructure is driving exponential increases in data movement, power consumption, and system complexity. As GPU cluster sizes climb into hundreds of thousands and rack power densities exceed 100 kW, the electrical interconnects that have served data centers for decades are running into hard physical limits on bandwidth, reach, and energy efficiency. MicroLED-based optical interconnects offer a path to sub-picojoule-per-bit energy efficiency, dramatically higher bandwidth density, and materially lower total cost of ownership — positioning optical I/O as a foundational, not optional, layer of next-generation AI infrastructure.

Table of Contents

1. Market Growth

AI progress is no longer constrained by how fast a chip can compute — it's constrained by how fast data can move between chips. Every frontier model today is trained not on a processor but on a network of tens of thousands of them, and the performance of that cluster is set by its slowest link: the interconnect. As GPU counts scale, electrical interconnects hit hard physical walls — bandwidth, power, distance, heat — and every hour a $40,000 accelerator sits idle waiting on data is capital burned.

Over the last 20 years peak compute on a chip has gone up about 60,000x, while the memory bandwidth to feed it has improved around only 100x, and the links between chips around 30x. The biggest bottleneck in AI now is the interconnect.

This is why the optical interconnect market is undergoing a structural re-rating driven almost entirely by AI infrastructure buildout. Independent research firms differ on absolute market size depending on scope — components versus modules versus systems — but agree directionally: the AI-driven segment is growing several times faster than the broader optical interconnect market and is rapidly becoming its dominant component.

1.1 Market Sizing Across Research Firms

Estimates of the total optical interconnect market by 2030 range from roughly $34B to $43B at a CAGR of 11–15%, reflecting the broad-based, multi-end-market nature of the category (telecom, enterprise, data center, AI):

Source	2025 Market Size	2030 Forecast	CAGR
Grand View Research	$17.85B	$34.54B	14.1%
AWS Marketplace / NMSC	$20.92B	$41.36B	14.6%
Fortune Business Insights	$15.38B	$43.14B (2034)	12.2%
Research & Markets	~$16.2B	n/a (2030 horizon)	11.2%

Sources: Grand View Research, Optical Interconnect Market Report (2025); Next Move Strategy Consulting / AWS Marketplace (2025); Fortune Business Insights, Optical Interconnect Market (2025); Research & Markets (2024).

1.2 The AI-Specific Segment Is Growing Far Faster

Within the broader category, the AI-specific slice is the standout growth story and is reshaping how the market is segmented and valued:

- Bank of America projects the AI optical connectivity market will grow from roughly $14B in 2025 to $73B by 2030 — a 39% CAGR — with optical ports rising from a minority share today to 71% of all AI network ports by 2030, as copper falls to 29%.
- TrendForce forecasts the co-packaged optics (CPO) and near-packaged optics (NPO) market specifically will expand from roughly $100M in 2025 to over $39B by 2030, with growth accelerating sharply between 2028 and 2029 as scale-up (chip-to-chip) optical architectures go into volume production.

- LightCounting forecasts 30–35% annual growth for AI optical transceivers in 2025–2026, moderating to 15–20% annual growth in 2027–2030 as the initial AI infrastructure wave matures.
- Optical interconnect content per AI server rack has grown from roughly $2,400 in 2020 to an estimated $14,000–$18,000 in 2025 — a structural, volume-independent growth driver as port counts and per-port speeds both rise.
- The chip- and board-level interconnect segment — the category most directly relevant to MicroLED optical I/O — is forecast to be the fastest-growing segment of the entire optical interconnect market, with a CAGR of roughly 15.7%, ahead of metro/long-haul and other legacy segments.

1.3 Competitive Landscape

The market is bifurcating into two tiers. Diversified incumbents — Broadcom, Marvell, Lumentum, Coherent Corp., Ciena, and NVIDIA itself — compete on integrated ASIC-plus-photonics platforms and direct hyperscaler relationships. A second tier of specialists and well-funded startups, including Ayar Labs (backed by Intel, NVIDIA, and GlobalFoundries), Lightmatter, and several silicon photonics pure-plays, are racing to commercialize chip-to-chip optical I/O — the segment most structurally aligned with MicroLED-based interconnects. Investment is concentrated in North America, China/East Asia, and Taiwan/Singapore, reflecting hyperscaler demand, national photonics initiatives, and manufacturing capacity respectively.

Kopin and its partner(s) like Fabric.Ai have an initial head start within the MicroLED space. Kopin has led the advanced MicroLED display market for several years due to the firm’s long aerospace and defense market presence which utilizes the brightness and contrast benefits of native LEDs. Further, Kopin invented and patented the bi-directional AI enabled microdisplay in 2022.

These technical achievements provide Kopin and Fabric.Ai with a significant market advantage in transceivers like Neural I/o™ and co-packaged optic transmitters.

2. Power as the Primary Constraint

AI systems are increasingly power-limited, not compute-limited. Every additional watt consumed moving data between GPUs is a watt unavailable for compute, and every watt of waste heat compounds cooling costs and rack-density limits. This dynamic is now the central design constraint across hyperscale AI infrastructure.

- AI training racks today routinely draw 80–120 kW, versus 10–15 kW for a typical enterprise rack a decade ago.
- Individual hyperscaler cluster buildouts are scaling toward gigawatt-class total facility power demand; some announced “AI campus” projects already plan for multi-gigawatt capacity.
- Interconnects — the links moving data between GPUs, switches, and storage — represent a rapidly growing share of total rack and cluster power, as compute itself becomes more efficient per FLOP while data movement volumes grow faster than compute.

2.1 Macro Energy Context

The scale of this challenge is corroborated by independent energy-sector analysis. The **International Energy Agency** projects that global data center electricity consumption will roughly double from 415 TWh in 2024 to approximately 945 TWh by 2030 — comparable to Japan's entire current electricity consumption — with AI-focused data center demand specifically tripling over the same period.

- Data centers are projected to drive more than 20% of total electricity demand growth in advanced economies through 2030.
- In the United States, data centers are expected to consume more electricity by 2030 than the country's combined energy-intensive manufacturing sectors — aluminum, steel, cement, and chemicals together.
- The U.S. and China together account for nearly 80% of projected global data center electricity growth to 2030.

Against this backdrop, interconnect energy efficiency is no longer a secondary engineering concern — it is a first-order lever on whether planned AI capacity can be powered, sited, and cooled on the timelines the industry is targeting.

3. How MicroLED Achieves Its Advantage: Speed, Parallelism, and Scalability

The efficiency and cost advantages described in this paper flow from a single architectural choice: MicroLED interconnects move data over **many slow channels instead of a few fast ones**. Understanding that choice — and being precise about what MicroLED is and is not — clarifies why the technology scales where copper and laser-based optics struggle.

3.1 Three Technologies, Three Approaches

Copper links move data as electrical signals over metal traces. Laser-based optics — whether pluggable modules or co-packaged optics — encode data onto laser light carried over fiber. MicroLED interconnects take a third path: arrays of microscopic LEDs, each switched on and off directly, each acting as its own independent data channel. The three approaches scale along different axes, and the axis determines where each one hits a wall.

3.2 Speed Per Lane: An Honest Comparison

On a per-lane basis, MicroLED is the **slowest** of the three technologies — not the fastest. A state-of-the-art copper lane carries roughly 224 Gbps; laser-based optical lanes run at roughly 100–200 Gbps; a single MicroLED channel carries on the order of 1–10 Gbps, with industry roadmaps targeting 20–25 Gbps.

MicroLED's advantage is not lane speed — it is **lane count**. A copper link is built from roughly 4–16 lanes, and a laser-based optical link from a handful to a few dozen channels, because each copper pair is physically bulky and each laser channel carries its own expensive light source and alignment hardware. A MicroLED array, by contrast, packs hundreds to more than a thousand emitters into a few square millimeters — industry demonstrations range from 128-channel links to arrays of 400+ emitters, and a 32×32 array alone provides 1,024 parallel channels. Each emitter is tiny, laser-free, and draws only microwatts.

The result: A MicroLED link can offer roughly **100× the parallel channels** of a copper link — a count comparison, not a speed comparison — and industry demonstrations have already aggregated those channels into links of one terabit per second and beyond. Because each channel runs slowly, it can also use the simplest possible signaling, skipping the power-hungry processing that fast serial lanes require. That simplicity is precisely where the sub-picojoule-per-bit efficiency described in Section 4 comes from.

Comparison of interconnect approaches (figures are approximate and configuration-dependent):

Technology	Speed per lane	Lanes per link	How bandwidth grows
Copper (electrical)	~224 Gbps (fastest)	~4–16	Push each lane faster every generation
Laser-based optics	~100–200 Gbps	~4–64	Faster lanes plus a modest number of channels
MicroLED optics	~1–10 Gbps (slowest)	Hundreds to 1,000+	Add more emitters — more channels, not faster ones

3.3 Scaling by Adding Pixels, Not Redesigning the Data Center

Parallelism also gives MicroLED its most distinctive scaling property. With copper and laser-based optics, more bandwidth means engineering a faster lane — new signaling standards, more complex electronics, and often new cables, connectors, and board designs to match. Each generation ripples outward into the surrounding infrastructure.

A MicroLED interconnect is designed to scale differently: because every emitter is an independent channel, capacity grows by populating the array more densely — **adding pixels, not re-engineering the link**. The additional bandwidth is expressed inside the same physical footprint, so the connector, the board layout, and the rack and cabling architecture around it can remain unchanged as capacity rises. In an industry where every change to data center topology carries enormous cost and delay, an interconnect designed to scale by channel count rather than by structural redesign is a materially different economic proposition — and it compounds the power, cooling, and capital advantages quantified in the sections that follow.

3.4 The Roadmap: Compounding Toward 20 Tb/s While Copper and Lasers Hit the Wall

The scaling path described above is not hypothetical — it is arithmetic. MicroLED bandwidth is the product of two independent variables: channels per array and speed per channel. Both are on published improvement curves, and neither requires a physics breakthrough. A 32×32 array already provides 1,024 parallel channels; at today's demonstrated per-channel rates, that yields roughly 1–10 Tb/s of aggregate bandwidth per link. As per-channel speeds advance along industry roadmaps toward 20–25 Gbps — still an order of magnitude slower than a copper lane, and therefore still simple, cool, and efficient — that same 1,024-channel array reaches 20–25 Tb/s. Denser arrays push the ceiling further: because emitters are microns across, arrays of 2,000 or 4,000 channels fit within the same few-square-millimeter footprint, implying aggregate capacities of 50 Tb/s and beyond without abandoning the slow-lane signaling that makes the technology efficient in the first place.

Copper and laser-based optics have no comparable path. Each scales on lane speed alone — a single multiplier — and lane speed is precisely where both are approaching hard physical limits: signal loss and power for copper, laser complexity and cost for optics. Every generation of improvement gets harder, hotter, and more expensive. MicroLED's roadmap runs in the opposite direction: each step adds channels that are individually simple and cheap, so capacity compounds while per-bit cost and power fall. The roadmap figures cited here reflect published industry projections for the MicroLED category and, like all roadmaps, carry execution risk — but the direction of the arithmetic is not in dispute.

Performance figures in this section reflect published third-party industry demonstrations and roadmaps (Avicena, Foxconn Research Institute, and industry conference disclosures), cited as evidence for the MicroLED category as a whole rather than as measured results for any specific product.

4. MicroLED Efficiency Advantage

MicroLED-based optical interconnects are designed to operate below 1 picojoule per bit (pJ/bit) of energy consumption, compared to 3–6 pJ/bit for conventional VCSEL-based optics and 5–10 pJ/bit for copper SerDes links at the reach and data rates relevant to AI fabrics.

4.1 Energy per Bit Comparison

Technology	Energy per Bit	Typical Reach
Copper SerDes	~5–10 pJ/bit	<2 m at 200G+
VCSEL / Silicon Photonics Optics	~3–6 pJ/bit	Meters to kilometers
MicroLED Optical (target)	<1 pJ/bit	Rack and pod scale

This represents:

- A 3–10× improvement versus copper SerDes.
- A roughly 2–6× improvement versus traditional VCSEL or silicon-photonic optics.

Physically, this matters because copper's reach is shrinking as data rates climb: at 200 Gbps per lane, high-quality copper cabling is effective only to about 2 meters before electrical losses (around 22 dB) make signals unrecoverable, rendering copper increasingly unviable even within a single rack. This is the central physics-driven force pushing the industry from copper toward optical I/O at progressively shorter reaches — a transition that BofA expects to push optical ports to 71% share of all AI network ports by 2030.

5. Power Savings at Scale

In large AI clusters, MicroLED interconnects can reduce interconnect-related power consumption by up to 75% relative to conventional optics, translating into tens of megawatts of savings at hyperscale.

5.1 Illustrative Model: 100,000-GPU Cluster

- Assumed interconnect bandwidth per GPU: ~1 Tbps
- Total cluster data movement: ~100 Pbps

Technology	Energy Intensity	Total Power
Traditional optics (~4 pJ/bit)	~400 kW per Pbps	~40 MW
MicroLED optics (~1 pJ/bit)	~100 kW per Pbps	~10 MW

Net modeled savings: ~30 MW per 100,000-GPU cluster.

To put 30 MW in context: it is roughly the continuous draw of a mid-sized industrial facility, or enough power for approximately 20,000–25,000 average U.S. homes — recovered purely from interconnect efficiency, without touching compute performance, on a single cluster. As hyperscalers deploy many such clusters in parallel — and as individual campuses scale toward gigawatt totals — this saved capacity compounds directly into deployable GPU headroom.

6. Total Cost of Ownership (TCO)

Interconnect power savings cascade into multiple layers of data center economics: direct energy costs, cooling overhead, capital deferral, and hardware utilization. Together these layers compound into a meaningfully different TCO profile for hyperscale AI deployments.

6.1 Direct Energy Cost Savings

At a blended industrial electricity rate of $0.08–$0.15 per kWh, a 30 MW reduction in continuous draw translates to approximately $21M–$40M in annual energy cost savings per 100,000-GPU-scale cluster.

6.2 Cooling Cost Reduction

Data center cooling overhead scales with IT load: industry rules of thumb put incremental cooling power at roughly 0.3–0.7 W per watt of compute/network load saved, depending on facility PUE and cooling architecture (air vs. liquid).

- Applied to the 30 MW interconnect savings, this implies an additional ~10–20 MW avoided in cooling power.
- That corresponds to a further ~$7M–$25M in annual savings, on top of direct energy savings.

6.3 Infrastructure Deferral and CapEx Avoidance

Lower power density per unit of bandwidth allows operators to fit more GPUs per rack and per facility footprint before hitting power or thermal ceilings. This defers or eliminates the need for:

- New utility substations and grid interconnection upgrades — increasingly the long-pole constraint as the IEA warns that up to 20% of planned data center projects risk delay due to grid connection bottlenecks.
- Additional liquid cooling infrastructure and associated facility retrofits.
- Premature facility expansion or new-site capital commitments.

Across a hyperscale buildout, CapEx avoidance of this kind can plausibly reach hundreds of millions of dollars, particularly where it allows a deployment to proceed within existing grid allocation rather than triggering a new, multi-year interconnection process.

6.4 Improved Hardware Utilization

Lower thermal loads correlate with higher component reliability: reduced failure rates, less unplanned downtime, and higher effective computing output per dollar of deployed hardware over the asset's operating life.

6.5 Aggregate TCO Impact

TCO Lever	Estimated Annual Impact (per hyperscale cluster)
Direct energy savings	$21M–$40M
Cooling cost reduction	$7M–$25M

TCO Lever	Estimated Annual Impact (per hyperscale cluster)
CapEx avoidance (amortized view)	Hundreds of millions, deployment-dependent
Combined OpEx impact	$30M–$65M annually

7. Societal Impact

Reducing AI infrastructure power consumption has implications well beyond any single operator's P&L: it affects grid capacity, emissions trajectories, and the accessibility of AI computing itself.

7.1 AI's Growing Energy Footprint

The IEA's central ("Base Case") projection sees global data center electricity consumption roughly doubling to ~945 TWh by 2030 — about 3% of global electricity demand, up from roughly 1.5% in 2024 — with AI-focused demand specifically tripling over the same window. Electricity consumption tied to AI accelerators is projected to grow around 30% annually in this scenario, versus roughly 9% for conventional servers. Longer-term, some industry estimates put AI's share of global electricity consumption in the mid-single digits to low double digits by the mid-2030s under more aggressive adoption scenarios.

Without efficiency gains at the interconnect layer, this growth trajectory risks:

- Straining regional electrical grids — the IEA already flags transmission and interconnection bottlenecks as a binding constraint on planned data center capacity.
- Increasing absolute carbon emissions in regions are still reliant on fossil generation to meet incremental load.
- Limiting AI access in grid-constrained or developing regions, where new gigawatt-scale loads are hardest to site.

By cutting interconnect energy intensity by 30–50% relative to conventional optics, and total data center power by an estimated 10–20% when cooling and density effects are included, MicroLED interconnects directly reduce the increment of new generation capacity required to support a given amount of AI compute — at gigawatt scale, equivalent to avoiding the output of multiple large power plants.

7.2 Enabling Broader AI Access

Lower TCO lowers the capital and operating bar for deploying AI infrastructure, extending feasible deployment beyond the small set of hyperscalers that can absorb gigawatt-scale power costs today to:

- Mid-sized enterprises building domain-specific or on-premises AI infrastructure — a segment some forecasts put at 14,000–18,000 enterprise data centers globally upgrading to 400G+ optical interconnects by 2030.
- Government and sovereign AI compute programs, including national initiatives in the UK, France, Germany, and India that are building hyperscaler-grade optical architectures at sovereign scale.
- Research institutions and universities running large-scale model training and scientific computing workloads.

7.3 Environmental Impact

Reduced interconnect and total data center energy consumption translates into lower associated CO_2 emissions, reduced water usage for evaporative cooling systems, and a smaller physical footprint per unit of delivered compute — each an increasingly material consideration as data center siting faces growing local and regulatory scrutiny over water and grid impact.

Conclusion

MicroLED optical interconnects are a foundational technology for scaling AI sustainably, economically, and efficiently. With the AI optical connectivity market projected to grow from roughly $14B in 2025 to as much as $73B by 2030, and global data center electricity demand on a path to nearly double over the same period, the case for interconnect efficiency has moved from a technical preference to a structural requirement.

This transition to optical interconnects is not incremental — it is structurally necessary, and represents the next evolution of semiconductors, electronic design, and AI computing.

MicroLED-based interconnects:

- Unlock Tbps-scale bandwidth density at the chip and rack level.
- Deliver step-function reductions in power consumption — sub-1-pJ/bit versus 3–10 pJ/bit for incumbent technologies.
- Enable economically viable hyperscale AI by directly reducing the OpEx, CapEx, and grid-interconnection burden of new capacity.

Most importantly, they reduce the cost of intelligence, lower the energy burden of computation, and enable a more sustainable and accessible AI-driven society — at a moment when global electricity demand from data centers is set to rise by over 500 TWh this decade, and when AI-specific demand is poised to triple.

Sources

[1] Grand View Research, “Optical Interconnect Market Size, Share & Trends Report, 2025–2030.”

[2] Next Move Strategy Consulting / AWS Marketplace, “Optical Interconnect Market Demand & Outlook – 2030.”

[3] Fortune Business Insights, “Optical Interconnect Market Size, Share & Trends [2034].”

[4] Research and Markets, “Optical Interconnect Market Size, Share & Forecast to 2030.”

[5] TrendForce, “Optical Interconnects Become Critical to AI Factory Expansion; CPO/NPO Market Expected to Exceed US$39 Billion by 2030” (June 2026).

[6] Bank of America / C-LIGHT, “AI Data Center Optical Transceiver Module Market 2025–2030.”

[7] DataM Intelligence, “Optical Interconnect in AI Data Centers Market Outlook 2033.”

[8] Dataintelo, “Optical Interconnect for AI Market Research Report 2034.”

[9] International Energy Agency, “Energy and AI” (2025) and “Key Questions on Energy and AI” (April 2026).

[10] S&P Global, “Global data center power demand to double by 2030 on AI surge: IEA” (2025).

[11] Data Center Dynamics, “IEA: Data center energy consumption set to double by 2030 to 945TWh” (2026).

[12] Brookings Institution, “Global energy demands within the AI regulatory landscape” (2026).

Note: Market sizing figures vary across research firms due to differing scope definitions (components vs. modules vs. systems) and segment boundaries. Figures are presented as reported by each source for directional and comparative purposes; this paper does not independently verify third-party market forecasts.