

The AI Infrastructure Barbell

Why the winners bet both ends of custom and open — and design for power and cooling first

By Mark McNiel · Business Development & Commercialization — AI Infrastructure, Optical Interconnect & Liquid Cooling · July 2026

Executive summary. As AI data centers scale toward million-accelerator clusters, operators face a defining architectural choice: build on custom, proprietary silicon, or on open, standardized (OCP) building blocks. Treated as an either/or, both paths carry hidden risk. This paper argues for a *barbell* — concentrate investment at both extremes and starve the fragile middle: deep custom co-design where a genuine edge justifies it, open standards everywhere nimbleness matters more. It also names the two risks the debate underplays — the power-and-cooling envelope and an accelerating consolidation wave — and offers a way to decide how to allocate between them.

1. The false choice

The prevailing conversation pits custom silicon against open standards as though an operator must pick a side and defend it. That framing is a trap. The right answer depends on three variables: the **use case**, the **scale**, and the operator's need to stay **nimble** as technology and market conditions shift. Teams that adopt one philosophy wholesale inherit its full downside — either the rigidity and cost of bespoke systems, or the performance ceiling of the lowest common denominator. The more useful question is not which architecture is superior, but how to hold both in the right proportion.

2. The barbell: bet the ends, starve the middle

Borrowed from risk management, a barbell concentrates resources at two opposite extremes and deliberately avoids the mushy middle. Applied to compute architecture, one end is **custom, co-designed silicon and interconnect** for the most extreme, highest-value workloads — where bespoke integration buys maximum performance, differentiation, and a durable moat. The other end is **open, standardized (OCP) building blocks** for everything else — where interoperability buys speed, reach, optionality, and second-source safety. The middle to avoid is the semi-custom “snowflake”: a one-off that carries custom cost and engineering burden without the strategic payoff, and never standardizes enough to scale. That middle is where budgets and roadmaps quietly go to die.

3. Where each architecture falls short

Each end has a real downside, and each is manageable only when it is hedged. **Custom and proprietary** designs lock the operator to a single roadmap, carry premium cost and vendor pricing power, and are slow to pivot when the workload or the underlying technology moves. **Open and standards-based** designs concede lowest-common-denominator performance, lag the bleeding edge because standards are written by committee, offer thinner differentiation, and risk fragmenting into competing camps that each claim to be “the” open standard. Neither downside is disqualifying — but either becomes fatal when an operator takes it on without a counterweight. The barbell is that counterweight.

4. The blind spot both share: power and liquid cooling

Whichever architecture an operator chooses, the binding constraint is no longer compute — it is **power delivery and heat removal**. Rack densities are pushing past 100–250 kW, and air cooling is finished at those levels; direct-to-chip liquid and immersion have moved from exotic to mandatory. Co-packaged optics — essential to scaling interconnect beyond the reach of copper — add heat-sensitive lasers to an already strained thermal budget. An architecture that outruns the facility's power and cooling envelope does not ship, no matter how elegant the silicon. The operators who win stop treating silicon, interconnect, power, and cooling as four separate procurement tracks and start **co-designing them as one system**. This is the discipline the custom-versus-open debate most often ignores, and it is where many otherwise-sound deployments stall.

5. The consolidation risk

The independent silicon field is consolidating rapidly. In roughly a year, NVIDIA acquired Groq, SoftBank absorbed Graphcore and Ampere, AMD took Untether's team, Intel moved on SambaNova, and Cerebras exited via public offering. For a buyer, betting a stack on a single proprietary vendor now carries explicit **M&A risk**: the roadmap can shift under you, pricing can move, or the technology can land inside a competitor. Open, standards-based building blocks are the hedge — when the field consolidates, portability preserves freedom of action. This is not an argument against custom silicon; it is an argument for not concentrating all of one's exposure in a vendor whose independence is no longer guaranteed.

6. Proprietary versus open is a risk decision

Reframed correctly, the choice is not a contest of engineering philosophies but an exercise in **risk allocation**. Proprietary buys performance and optimization while concentrating risk: lock-in, pricing exposure, and stranded-asset risk if the vendor pivots or is acquired. Open buys optionality and resilience while conceding the bleeding edge. The barbell resolves the tension by assigning each its proper role — take proprietary risk only where a genuine, defensible edge justifies it, and default to open everywhere continuity and nimbleness matter more. The governing rule is simple: never pay custom prices for commodity outcomes, and never accept commodity performance where you need to win.

Bet the ends. Starve the middle. Solve power and cooling first.

Conclusion

The AI buildout rewards operators who refuse the false binary. Bet the ends — deep custom where it earns its keep, open standards everywhere else — and starve the fragile middle that offers the advantages of neither. Weigh proprietary and open as a risk decision, not a loyalty test, and hedge the consolidation wave with portability. Above all, design for power and cooling first, because the most advanced silicon in the world is worthless if the facility cannot feed and cool it. The winners will not be the operators with the most exotic chips; they will be the ones who assembled silicon, interconnect, power, and cooling into a single, deliberate, and adaptable system.

Mark McNiel — AI Infrastructure, Optical Interconnect & Liquid Cooling. Views are the author's own. Company references reflect publicly reported activity as of July 2026; forward-looking items (e.g., SambaNova–Intel) are reported or in progress, not closed.