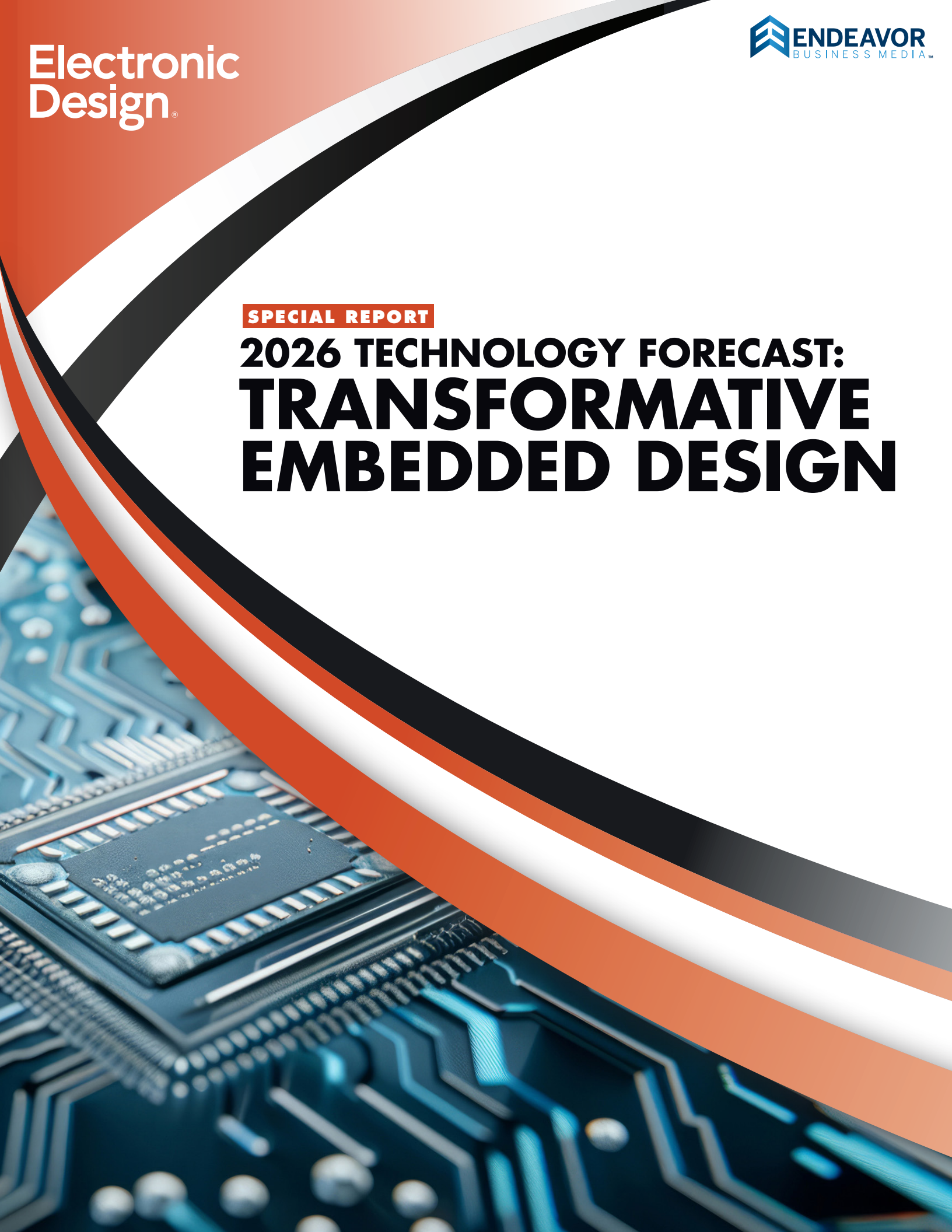
Electronic Design

ENDEAVOR BUSINESS MEDIA

SPECIAL REPORT

2026 TECHNOLOGY FORECAST: TRANSFORMATIVE EMBEDDED DESIGN



SPECIAL REPORT

Table of Contents

2026 Technology Forecast: Transformative Embedded Design

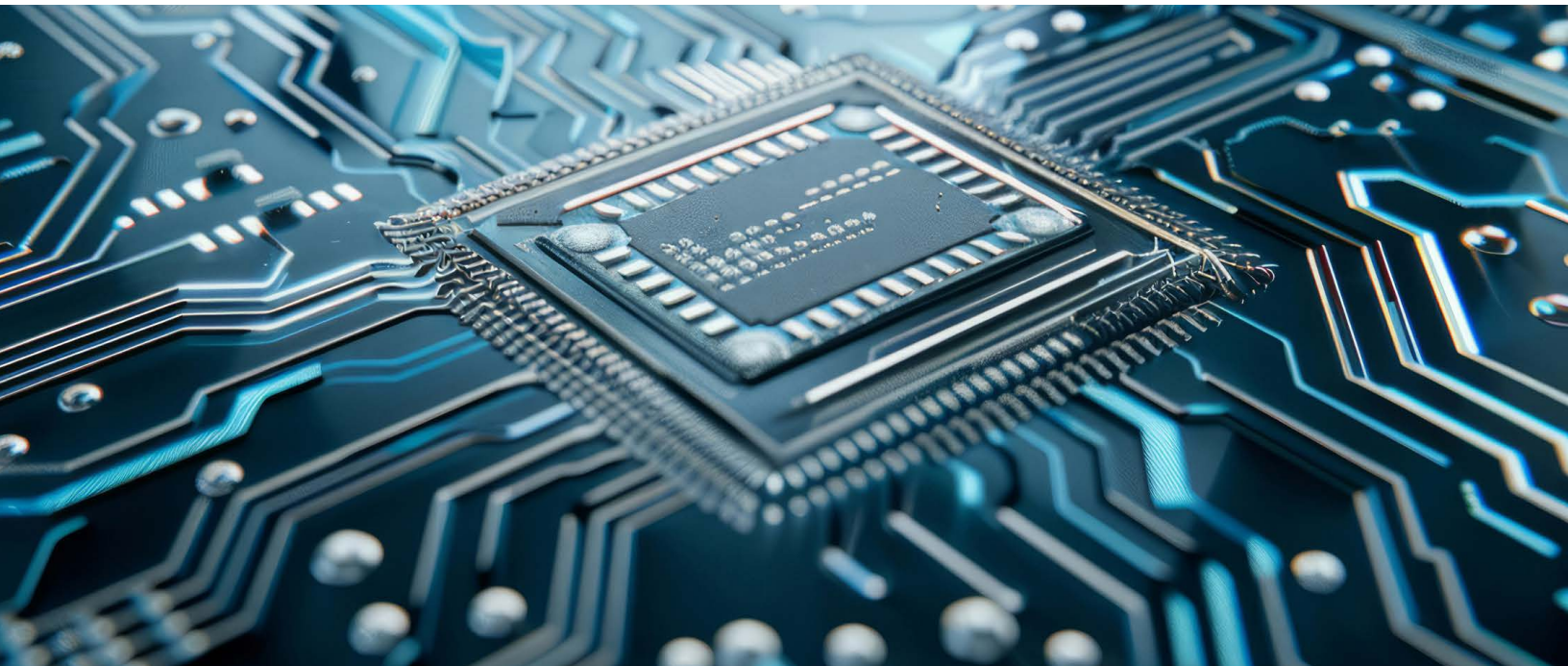
William G. Wong,
Senior Content Director
Electronic Design & Microwaves & RF

TECHNOLOGY CONTINUES to advance even in the face of supply chain issues due to war and tariffs. This year we asked our contributors to focus on a single topic area to get more in-depth insights into where those technologies are headed.

This Special Report addresses a range of topics important to embedded developers from how hyperscalers are dominating the use of, as well as advances in, memory and processors, to where the ubiquitous Ethernet is headed.

We covered a lot of ground with the Chiplet Summit and the Special Report features a forward-looking article dealing with chip packaging and chiplets. This includes artificial-intelligence (AI) processors and where they're headed, too.

Finally, we look at Rust. This programming language delivers safe and secure features that aren't found in programming langauges like C, which have been the mainstay for embedded systems. Rust is even being used to support Linux development.



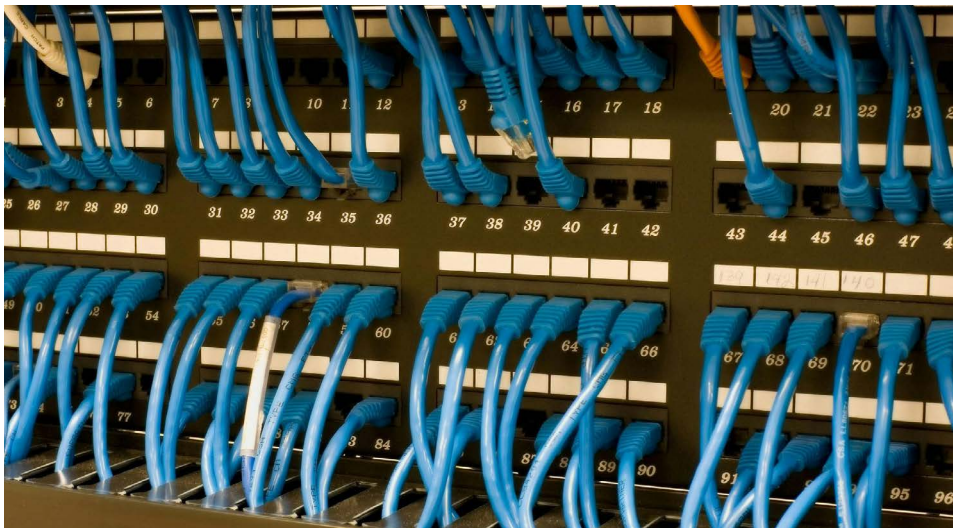
CHAPTER 1

2026 Ethernet Forecast: Full Speed Ahead

AI continues to be a driving force for Ethernet developments. This article forecasts the opportunities and challenges Ethernet technology will face in 2026 to keep up with the AI demand.

DAVID J. RODGERS, Ethernet Alliance President and Events & Conferences Chair, *Ethernet Alliance*

PETER JONES, Chair, *Ethernet Alliance*



Dreamstime_Nikm_762111

In 2026, Ethernet's trajectory will be influenced by AI, rapid progress in standards and signaling technology, and its continued move into markets and applications beyond the data center. The convergence of these three factors suggests a time of expansion and innovation unlike any seen before. What follows explores how these forces will shape Ethernet throughout 2026.

Escalating Ethernet

It should come as no surprise that Ethernet development continues to accelerate at an unheard-of rate. Even though IEEE 802.3 is still refining 200G/lane signaling specifications, consensus building around 400G/lane is well underway. What once felt like long-term planning is now happening in parallel.

With this exponential advancement, Ethernet has overtaken other interconnect technologies like Infiniband, making it the leader in "scale-out" AI networking. Connecting key AI functional blocks demands reliable, scalable, and interoperable products sourced from a broad and competitive market. Luckily, that openness and interoperability are at the heart of Ethernet's value proposition.

But the high-profile, high-speed hyperscale market is only the tip of the Ethernet iceberg. Momentum is building behind widely supported initiatives such as UA Link, Ultra Ethernet, and OCP's SUE-T and ESUN projects, fueling the world's unrelenting appetite for faster speeds. Add in the phenomenal proliferation of IoT devices — each with its own wired or wireless connection — across enterprise, industrial, and edge environments, and you've got the urgent need for an even more stable, predictable, and resilient Ethernet ecosystem.

Despite the swift pace of Ethernet standards development, the hyperscale market is moving forward in advance of ratified global standards. As revealed during the Ethernet Alliance's recent *Technology Exploration Forum 2025 (TEF 2025): Ethernet for AI*, there's a near-insatiable demand for high-speed, low-latency interconnects that can support booming AI and ML compute workloads.

Hyperscalers will persist in active-

ly pursuing customized solutions for intra-data-center optimization. However, interconnects spanning disparate physical locations will continue to rely on a robust catalog of stable, interoperable standards-based solutions.

The Ethernet Alliance 2026 Ethernet Roadmap illustrates the balance between near-term innovation and long-term interoperability. Instead of focusing on a single outcome, it captures how the ecosystem is seeking continuous improvement in speed, reach, power efficiency, and deployment models while maintaining broad compatibility. The roadmap underscores Ethernet's evolution not just at the hyperscale core, but across the full range of applications that increasingly depend on it.

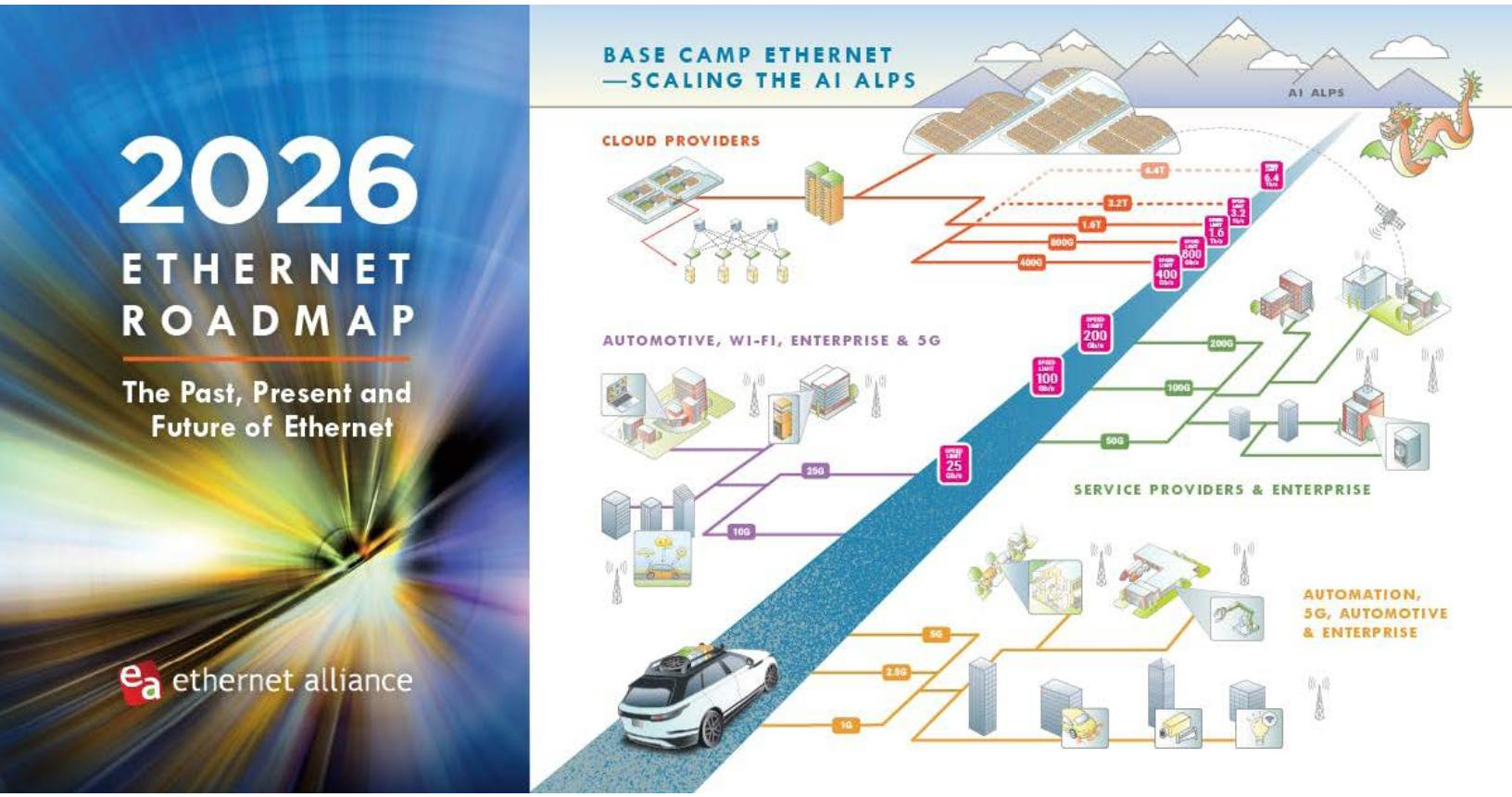
Scaling for What Comes Next

The 2026 Ethernet forecast is simple: AI, more AI, and even more AI.

The constraints shaping Ethernet's next phase aren't theoretical. They're physical, financial, and logistical:

- AI data centers and AI CPU orders are now being measured in gigawatts of power consumption.
- Leading semiconductor manufacturer TSMC estimates it can meet only a fraction of current demand for advanced-process devices.
- Wall Street thinks AI-driven hyperscaler capital spending may exceed $520B in 2026.
- The 10th most populous U.S. state, Michigan, just approved a contract to provide up to 1.4 GW for a single data center.
- Large power transformers required to support data center power demand are facing lead times exceeding two years.

Against that backdrop, Ethernet's presence in data centers continues to swell, with switch orders doubling



1. The past, present, and future of Ethernet speeds depicted in the Ethernet Alliance's 2026 Ethernet Roadmap. Ethernet Alliance

CHAPTER 1: 2026 Ethernet Forecast: Full Speed Ahead

between 2022 and 2025. Ethernet has moved ahead of InfiniBand in AI back-end networking, a trend that looks to be unstoppable at this point. Vendors are increasingly targeting scale-up networks previously dominated by proprietary interconnects, while 1.6-Tb/s switches are expected to ship in volume in 2026. And this may even end up being the year that co-packaged optics (finally!) gain ground.

On the standards front, IEEE 802.3 expects to complete IEEE 802.3dj, covering 200 Gb/s, 400 Gb/s, 800 Gb/s, and 1.6 Tb/s, by late 2026. At the same time, a 400-Gb/s/lane Signaling Call For Interest (CFI) is already scheduled for March.

There's active debate surrounding modulation approaches, but assuming PAM-6 for the moment and ignoring coding overhead, the math quickly becomes sobering. PAM-6 provides roughly 2.5 bits per symbol, implying symbol rates on the order of 160 GHz for 400G per lane. My friends in the SerDes and connector space tell me that at a wavelength of 0.12 mm, board and component design can get challenging pretty fast.

In other news, Ethernet's growth remains broad-based:

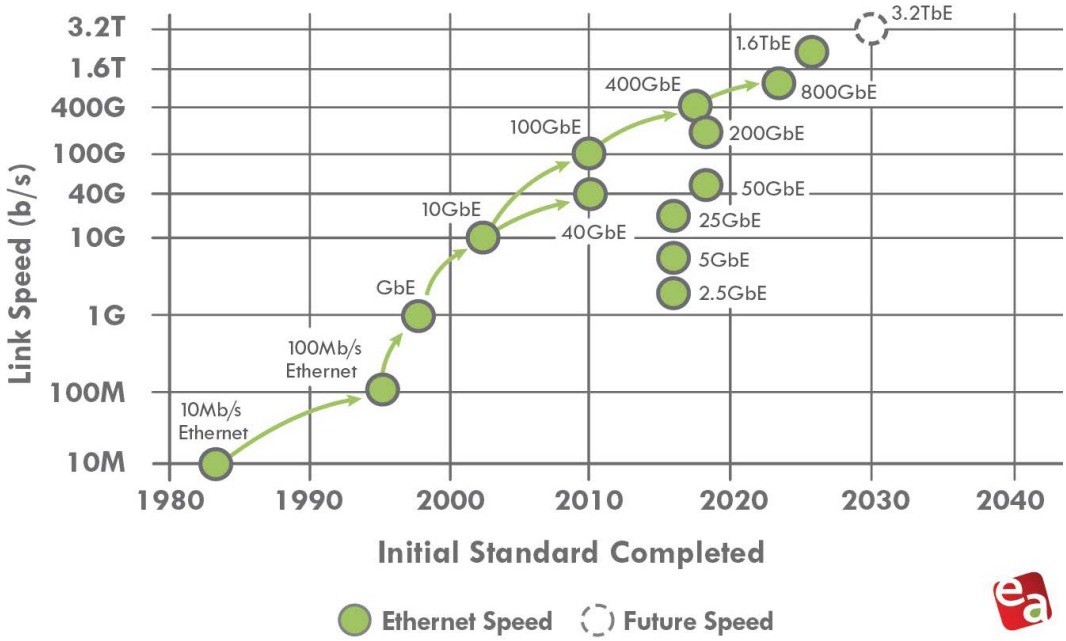
- The enterprise switch market is expected to exceed $30B in 2026.
- The industrial Ethernet market is expected to surpass $12B in 2025, with compound growth greater than 7% through 2032.
- Automotive Ethernet is projected to easily cruise on by the $3.3B mark in 2025 to $29.4B and a CAGR of more than 24% by 2035.

So, what's the bottom line? Ethernet is on the cusp of unprecedented growth, thanks especially to AI data center networks. Simultaneously, it remains the linchpin of enterprise, mobile, and home networking, with steady — if not vigorous — expansion across all three. It's also extending laterally into major segments such as industrial automation and automotive systems, while taking flight in emerging markets and applications, such as drones and spacecraft.

Never bet against Ethernet.

view this article online

BACK TO TABLE OF CONTENTS



2. The Ethernet Alliance's 2026 Ethernet Roadmap details Ethernet's expanding role in AI, automotive, enterprise, cloud, manufacturing, and next-gen connectivity. Ethernet Alliance

CHAPTER 2

Empower the Next Wave of Semiconductor Reuse Through Chiplet Realization

Unlock the future of semiconductor innovation with chiplets — enabling scalable, cost-effective designs and paving the way for a reusable chip marketplace.

MICK POSNER, Senior Product Marketing Group Director, Chiplet and IP Solutions, *Cadence Design Systems*

Shown are low-end, mid-range, and high-end chiplet systems (representative architectural diagrams, not drawn to scale).

Chiplets are clearly on the minds of system-on-chip (SoC) and electronic systems designers. Case in point: Cadence's eBook titled "Cadence Chiplet Solutions — Helping You Realize Your Chiplet Ambitions" was downloaded more than 500 times in its first month.

Chiplets enable a marketplace of reusable semiconductor die, allowing system designers to aggregate complex functions through package-level scaling. The resulting chips deliver immense functionality with unprecedented time-to-market, at a fraction of the typical cost for a monolithic chip design.

They're predicted to enable massive semiconductor reuse, assuming a large-scale chiplet marketplace is realized. The question isn't whether such a marketplace will materialize, but when it will materialize and what it will take for that to happen.

To understand the future of chiplets, it's worth pausing and touching on the user profile for multi-die designs and chiplets. Multi-die design is an established semiconductor design technique spearheaded by the data center, cloud infrastructure, and AI/high-performance computing (HPC) markets.

In this model, users typically design all die in the system, managing everything from design and manufacture to packaging. They sometimes utilize proprietary die-to-die interfaces or customize them to specific needs, treating die reuse as a nice-to-have, rather than a must-have.

On the other hand, chiplet users are unlikely to design all dies in the system. Instead, they're looking to outsource some or all of them, relying on standardization across die-to-die interfaces and packaging to ensure interoperability. These chiplet aggregators are likely planning multiple product SKUs from the same set of die and expecting future die reuse across numerous product generations.

Many applications utilize multi-die technology, with mature usage seen across data centers, cloud, and AI. The use of chiplet-based architectures across these applications continues to expand, primarily for I/O disaggregation, with several off-the-shelf chiplet solutions available for designers. As the importance and complexity of I/O

and memory interface designs ramps up, designers seek ways to accelerate design schedules without the need to build deep in-house expertise.

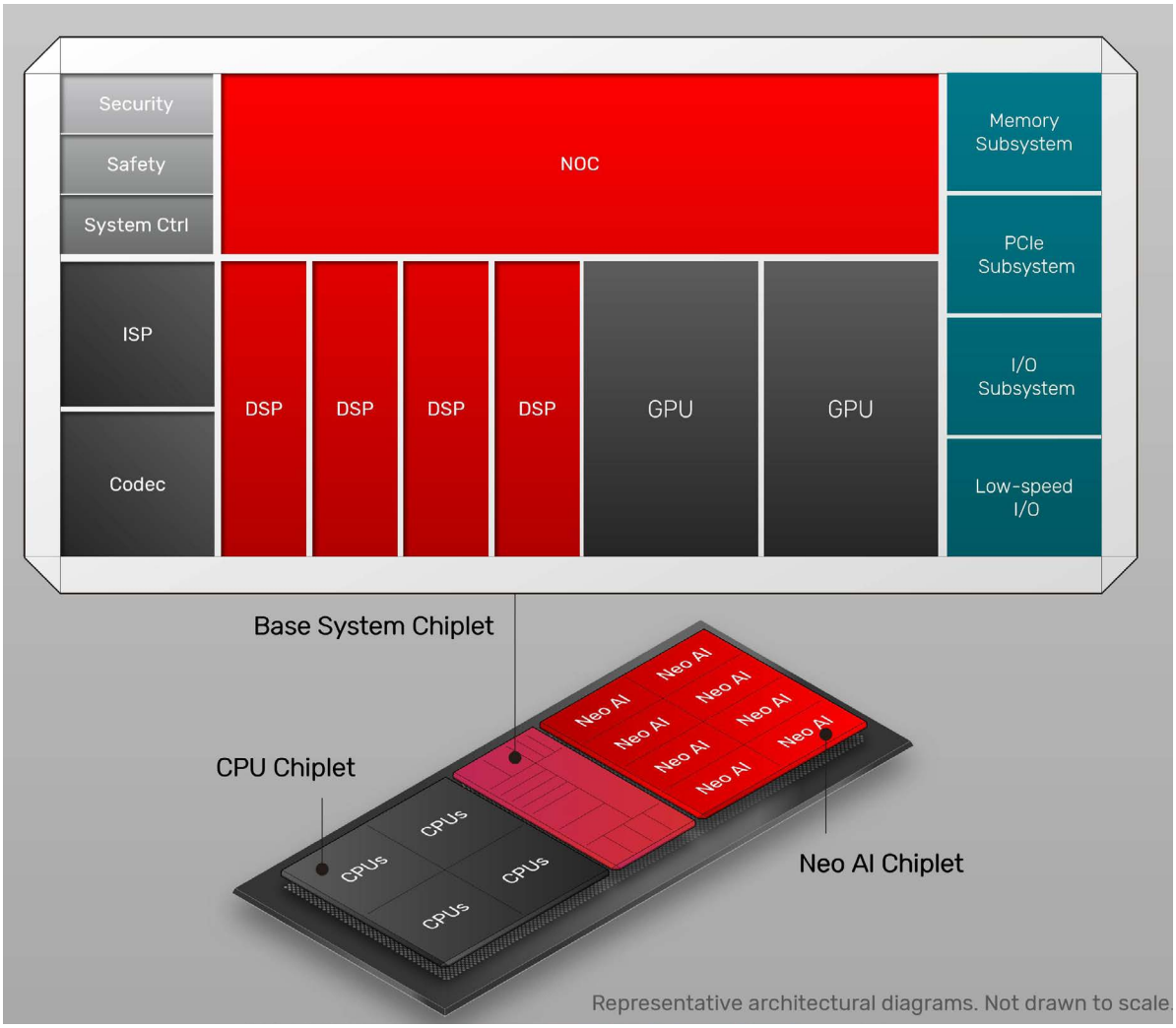
Enter Physical AI

Physical AI — the combination of automotive, robotics, drones, and aerospace and defense — looks to be the next big adopter of chiplets, which provide an effective path to enable scalable edge-AI-capable devices. Acquiring a silicon-proven chiplet die for these essential functions and package integrating, rather than designing from the ground up, achieves these goals.

With physical AI system silicon, the main system functions neatly disaggregate into logical partitions such as CPU processing, AI engines, and system utilities (I/O, memory). These die building blocks make possible a flexible chiplet system architecture covering a multitude of designs, including autonomous systems, object detection, natural language processing (NLP) and interpretation, and physical action applications **(Fig. 1).**

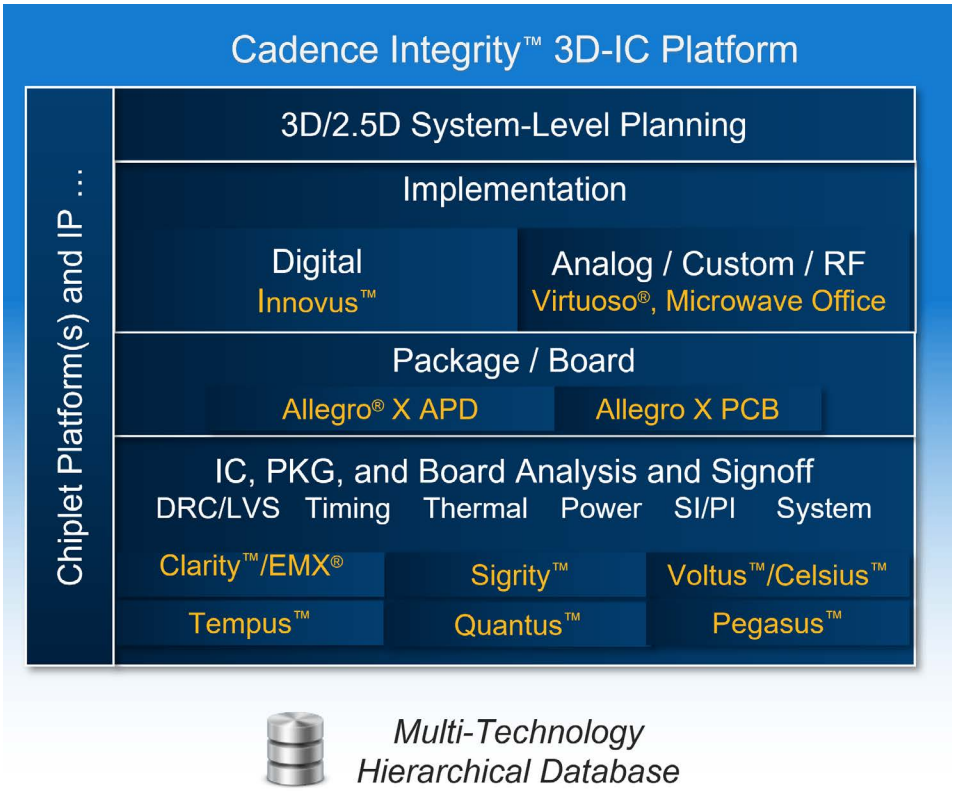
With the value proposition apparent, what's holding back widescale marketplace evolution and chiplet availability? Several factors play into this, including the fear of being first, standardization convergence, and a bit of a technology gap, all of which are solvable in my opinion.

Some companies don't have the time or money to risk adopting a relatively unproven chiplet system solution. Instead, they're waiting for silicon proof. In addition, these companies



1. This is a typical SoC to chiplet decomposition of a physical AI system.

CHAPTER 2: **Empower the Next Wave of Semiconductor Reuse Through Chiplet Realization**



2. Cadence Integrity 3D-IC EDA, IP, and Chiplet Platform is an example of EDA tools designed to support chiplet and chiplet-based system design.

don't want to fund such a development themselves, preferring to let someone else assume the development cost and prove the system. Only then will they adopt a similar architecture that matches their needs.

Comprehensive Specifications Accelerate Adoption

With the introduction of more comprehensive and inclusive specifications, much progress on chiplet system architecture standardization has been made in the past few years. However, many items are still to be determined. Competing specifications remain, too, and adoption hasn't yet driven convergence on a single standard over another.

Huge technology innovations have emerged over the years, including the availability of EDA tools that more seamlessly handle the increased complexities of chiplet-based design and verification **(Fig. 2)**. Packaging technologies also continue to grow in capabilities and maturity, yet interoperability between package types remains a challenge.

Looking ahead, expect the challenge gaps to close significantly, but not fully. (Hey, we have to be realistic — this is a journey, not a milestone.) Companies that fund reference platforms will pave the way for lower-risk adoption of chiplet-based architectures. Chiplet system architecture specifications that promote open and CPU-neutral interfaces will accelerate the convergence of standardization, setting the stage for individual chiplet providers to build interoperable solutions that feed into a chiplet marketplace.

Combined with a focus on building reference platforms, this will lead to innovative chiplet architecture frameworks that simplify standards-based chiplet design. Continued EDA and IP advances will reduce the complexity adoption barrier for chiplet-based designs by providing increased automation and repeatable flows.

The industry is on the verge of realizing the full potential of chiplets, which is to enable the aggregation of complex functions through package-level scaling. Looking ahead, we can expect to see the industry continue streamlining chiplet-based design and unlocking its immense potential. As for when that chiplet marketplace will emerge and facilitate large-scale adoption, time will tell, but the essential groundwork will be laid in 2026.

view this article online

BACK TO TABLE OF CONTENTS

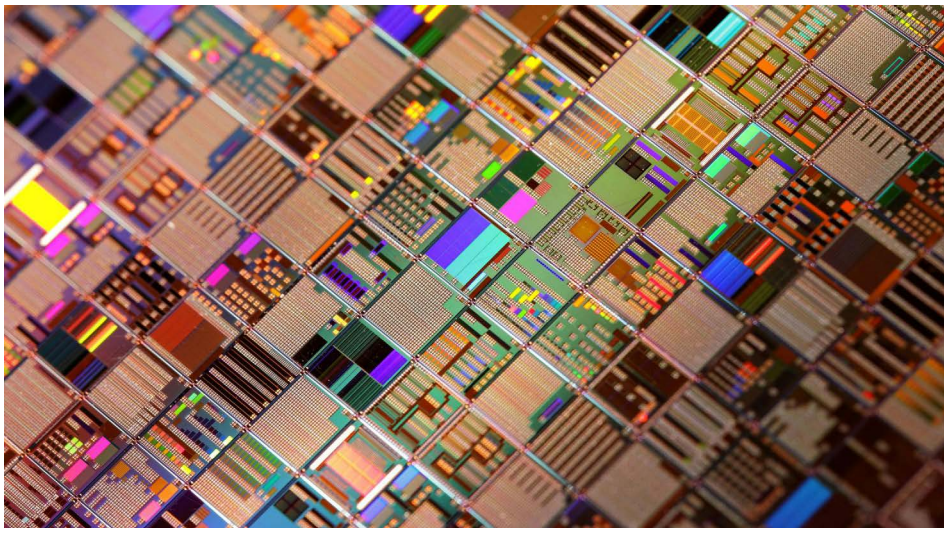
CHAPTER 3

Is It Make-or-Break Time for AI Processor Startups?

Mapping the fast-growing market for AI processors indicates the startup boom has peaked — or is very close to it.

JON PEDDIE, Persident, *Jon Peddie Research*

JAMES MORRA, Senior Editor, *Electronic Design*



Dreamstime_Andreyi-Armiagov_397851113

The explosive growth of NVIDIA and the insatiable demand for its GPUs have fueled a global surge in AI processors. However, the wave of startups building dedicated AI silicon has crested, or it's very close to it.

The number of AI processor (AIP) startups around the world has more than doubled since 2016, and by the end of 2025, the number of standalone companies in the area jumped to an unsustainable total of 146. An astounding sum of $28 billion has been invested in these companies to date, with investors lured by the promises of the AIP market. Estimates are that the market will top $494 billion in 2026, driven mainly by inference (cloud and local) and edge deployments (wearables to PCs) for units, and training and hyperscalers for dollars.

That's despite NVIDIA's technology being backed up by a deep, wide, and all-but-unassailable software stack and a total hardware infrastructure for a data center. None of that seemed to sink in for investors, who have been ready and willing to invest in almost anyone who said they could build a faster, smarter, cheaper AI processor.

As might be expected, most companies are focusing on AI inference either in data centers or out at the edge. Training remains very capital-intensive, and most startups have ceded the space to NVIDIA.

"The population of standalone AI processor suppliers will shrink by 40% over the next year or two, and the reality could be worse than that." — Jon Peddie Research

However, the window for most of these startups to succeed is likely closing. The peak formation year was 2018 (by then 75% of startups had already appeared). Interestingly, the rise in

"The population of standalone AI processor suppliers will shrink by 40% over the next year or two, and the reality could be worse than that." — Jon Peddie Research

the number of startups began before NVIDIA's business started booming, stunning the tech industry.

One would think that NVIDIA's success was the honey that attracted all of the ants, but a good number (58%) of the startup companies got going before that. Since 2022, the sector has averaged seven acquisitions per year, and since 2020, 17 of the startups have filled an IPO.

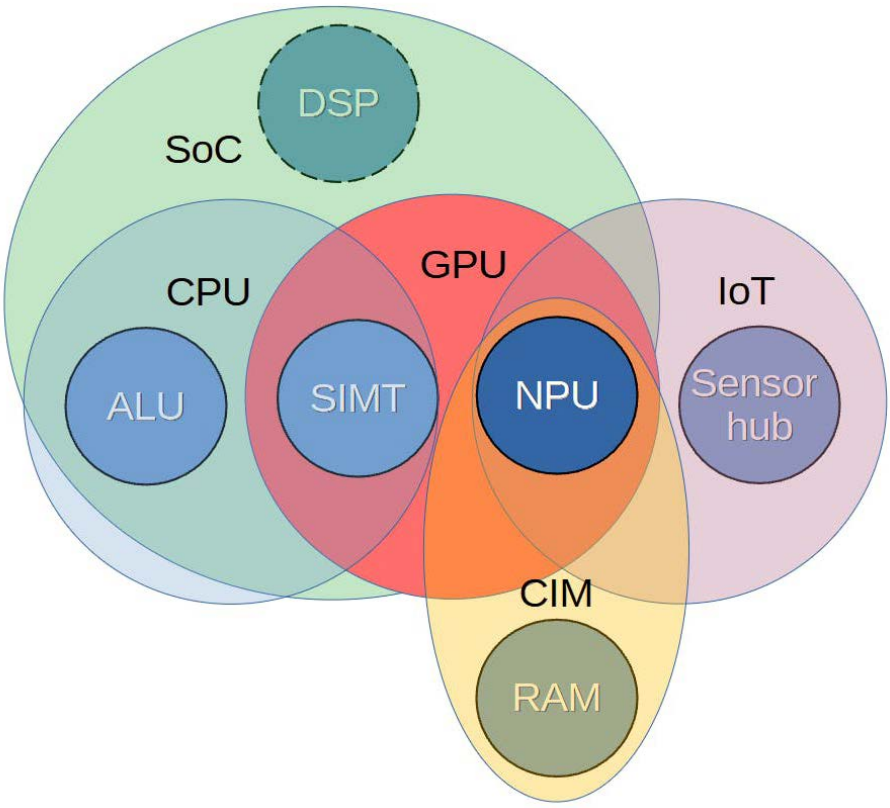
The Complex Landscape of AI Chip Startups

At a basic level, an AIP is a chip optimized to run neural-network workloads fast and efficiently by doing huge amounts of tensor math while moving data as little as possible. The processor types of offerings span GPUs, NPUs, compute-in-memory/processing-in-memory (CIM/PIM), neuromorphic processors, and matrix/tensor engines.

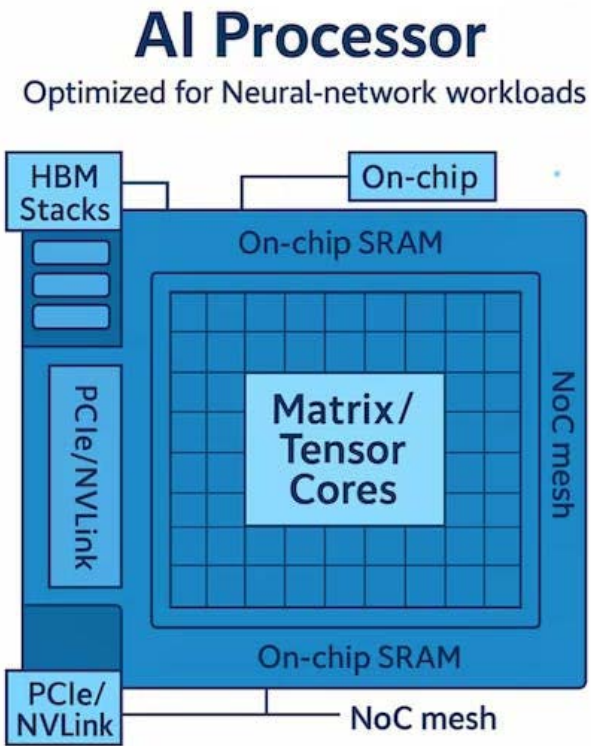
While CPUs and FPGAs are also used to run AI workloads, they tend to be evaluated separately from the $85 billion AI chip market since their generality makes it impossible to differentiate by function. But CPUs with vector extensions or SIMD engines (which is just about all of them) do fit in the AIP category as well. The overlap of CPUs, SoCs, and ASICs can be bewildering (**Fig. 1**).

At a fundamental level, AI processors come with several core building blocks (**Fig. 2**):

- **Compute blocks:** Wide SIMD/SIMT cores (GPU style), tensor or matrix



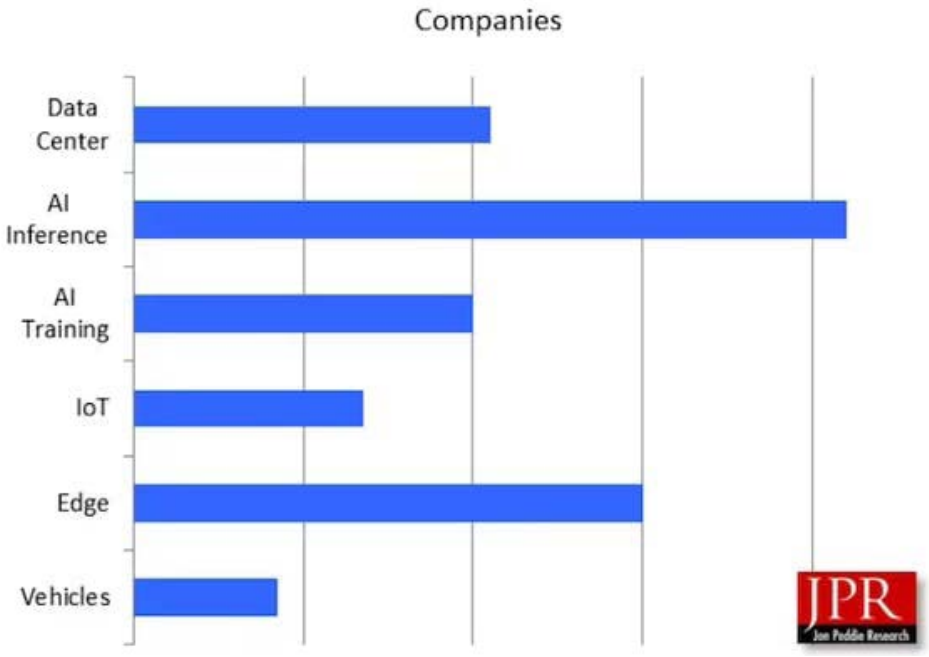
1. The complicated overlap of CPUs, SoCs, and ASICs used for AI training and inference. Jon Peddie Research



2. The building blocks of a typical AI chip, including compute, memory, and connectivity. Jon Peddie Research

engines, (NPUs) vector units, activation units.

- **Memory hierarchy:** Small, fast on-chip SRAM close to the compute; larger HBM/DDR located outside the processor or in the same package; caches or scratchpads; prefetchers/DMA (CIMs loosely fit in this category).
- **Interconnects:** On-chip network-on-chip (NOC) and off-chip links, including (but not limited to) PCIe, CXL, NVLink, and Ethernet.
- **Control:** Command processors, schedulers, and microcode for kernels/collectives.



3. Distribution of AI chip companies by market segment. Jon Peddie Research

The world of AI processors spans cloud services, data center chips, embedded IP, and neuromorphic hardware. Founders and engineers address the gaps that CPUs and GPUs can't fill: managing memory; maintaining high utilization with small batches; meeting latency goals on strict power budgets; and providing consistent throughput at scale.

Companies develop products along two main dimensions: the type of workload — training, inference, or sensor-level signal processing — and the deployment tier, from hyperscale data centers to battery-powered and wearable devices.

Most technical work has centered on memory and execution control. CIM and analog techniques reduce data transfers by performing calculations within memory arrays and keeping partial sums nearby, which can lead to data-flow designs. Wafer-scale chips store activations in local SRAM and stream weights for long sequences.

Reconfigurable fabrics alter data flow and tiling during compile time to optimize utilization across multiple layers. And training chips emphasize interconnect bandwidth and collective communication, while inference chips prioritize batch-one latency, key-value caching for transformers, and power efficiency at the edge, as well as cloud independence to reduce latency (especially important in agentic robots).

Adoption depends on go-to-market strategies and ecosystem backing. Cloud providers are incorporating accelerators into managed services and model-serving frameworks. IP vendors are working with handset, auto, and industrial SoC teams, offering toolchains, models, and density roadmaps.

In addition, edge specialists release SDKs that compress models, quantize to INT8 or lower, and map operators onto sparse or analog units while achieving accuracy targets. And neuromorphic groups put out compilers for spiking networks and emphasize energy efficiency and latency on event streams. Refinements in compilers, kernel sets, and observability tools often outweigh peak TOPS.

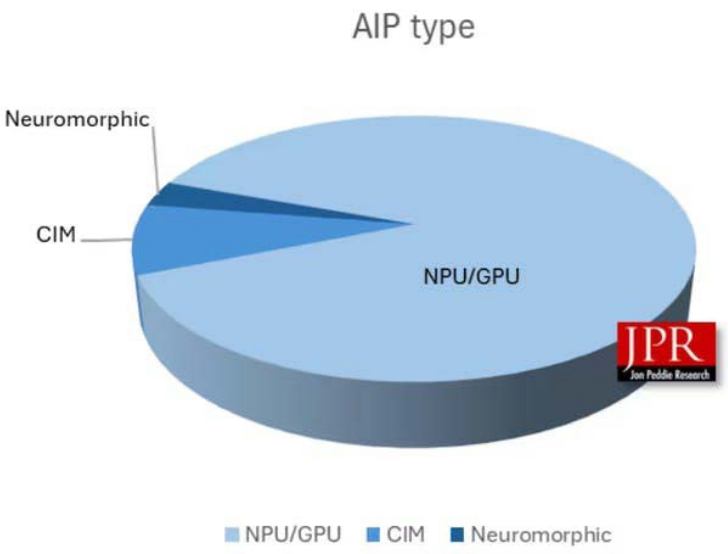
Competition varies by tier (**Fig. 3**). Training silicon focuses on cost per model trained, considering network, memory, and compiler constraints. Inference silicon targets cost per token or frame within latency limits, using cache management and quantization as tools. Edge devices compete on

CHAPTER 3: Is It Make-or-Break Time for AI Processor Startups?

milliwatts per inference and toolchain portability. IP vendors compete on tape-out time; power, performance, and area (PPA) goals; and verification support. Research projects balance speed to market against experiments that may alter the tradeoffs between memory, computing, and communication.

Throughout this process, teams customize designs to meet specific needs, such as attention depth, parameter count, activation size, sparsity, and precision policies. When companies synchronize silicon, compiler, and deployment tools, they reduce integration costs and speed up the transition from models to high throughput.

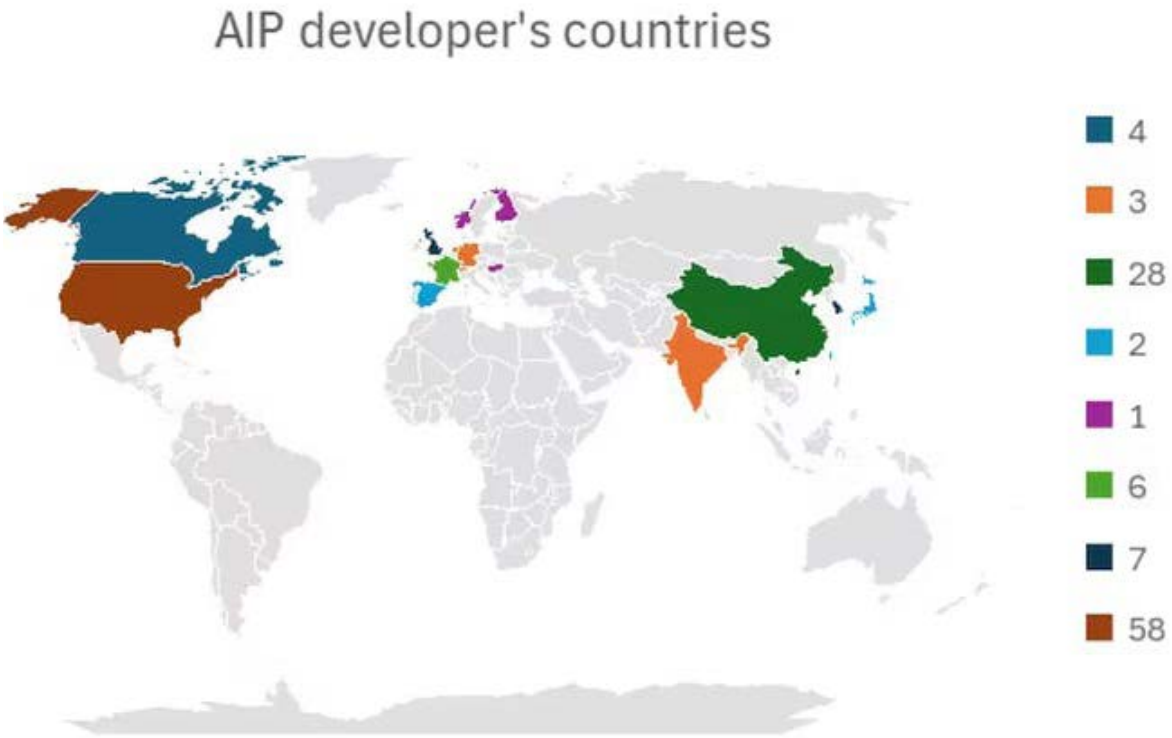
Customers then have multiple options: expand in the cloud, scale up with wafer-scale systems, embed NPUs in SoCs, or put computing closer to sensors using analog and neuromorphic chips. This vast volume of research and work is where the $28 billion investment is going, and what makes the startups attractive acquisition candidates (**Fig. 4**).



4. Distribution of AI chip startups based on processor type.
Jon Peddie Research

The End of the Explosive Growth of AI Chip Startups

But this "Cambrian explosion" in AI chip startups is likely coming to an end. In late 2025, the first signs surfaced of investors getting nervous about the



5. The global population of AI processor suppliers.
Jon Peddie Research

huge capital equipment expenditures being made by hyperscalers, governments, and private firms, causing ripples in a stock market largely driven by the AI boom (**Fig. 5**). Talk about a bubble market and forecasts as to when it would pop filled the press.

Brewing just below the surface is the inevitable bubble pop in AI silicon providers. After all, no industry can sustain 146 suppliers.

There have already been several acquisitions and failures (21 by the end of 2025) and more will come. But VCs bet on the odds, and the odds are that the six companies each with more than $1 billion in funding will be the survivors, and the other 100-plus startups will be looking for a home in one of the 37 public companies in acquisition mode. The crystal ball at Jon Peddie Research suggests the population of standalone AI chip suppliers will shrink by 40% over the next year or two, and the reality could be worse than that.

Even though most of the startups will be acquired or fail, those that are acquired will bring a bonus of free IP with them, paid for by the enthusiastic and optimistic VCs. You can buy a lot of R&D with $28 billion, especially when the average employee count of the startups is less than 10 people. For reference, NVIDIA currently has around 36,000 employees, so it's not exactly a fair fight.

These acquisitions will come with some broken hearts about what could have been. However, it's not merited if the premise of the dream was, "Build a better processor and the world will beat a path to your door."

While NVIDIA's dominance is backed up by the performance of its AI GPUs and deep software stack, its focus on building a total hardware infrastructure for a data center is keeping it ahead of the pack.

view this article online

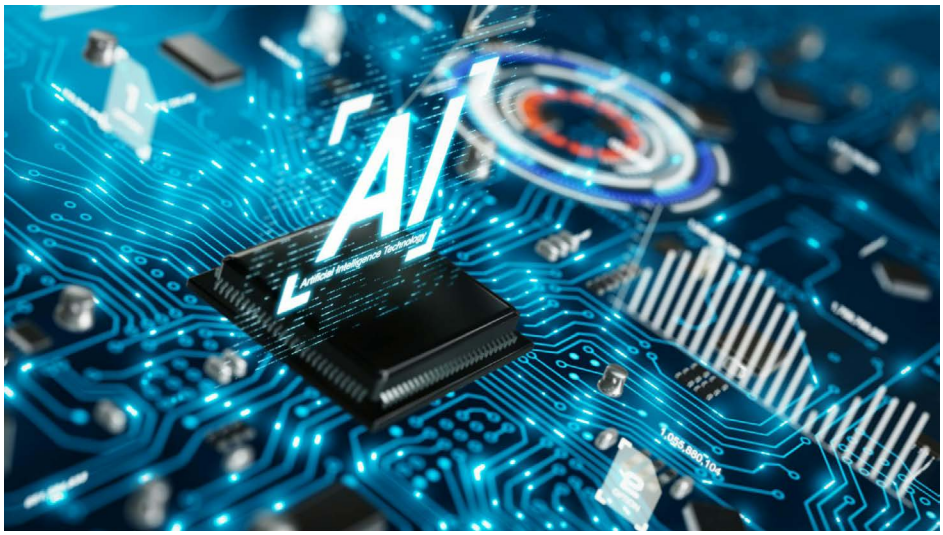
BACK TO TABLE OF CONTENTS

CHAPTER 4

What Happens When Hyperscalers Take Their Foot Off the Accelerator?

Hyperscaler spending currently paints a rosy picture for the memory and processor markets, but gloom looms in the future.

JIM HANDY, General Director, *Objective Analysis*



ID 137795258 © Thekaikoro | Dreamstime.com

The chip market is in an amazing place. All markets are growing, but two really stand out: memories and processors. But not just ANY memories and processors. DRAM, especially HBM DRAM, has been on a tear since 2024, which has created shortages of other types of DRAM, driving the prices up.

Meanwhile, NAND flash is beginning to enjoy a growing revenue surge that could lead to the market imitating DRAM and going into a significant shortage. As for processors, those that are doing best are GPUs and other AI-related processors like Google's TPU, with server CPUs following along.

AI-Centric Markets Are Driving Record Growth

While other semiconductors are enjoying the kind of growth they had before the pandemic, GPUs, DRAM, and NAND flash are setting historic records. The difference between the growth of AI and non-AI semiconductors in 2024 is dramatically evident in **Figure 1**.

Why is that? It's because hyperscalers are each trying to out-spend all others on AI. These companies have an enormous fear of missing out, and that's led to the explosive growth of hyperscaler capital spending.

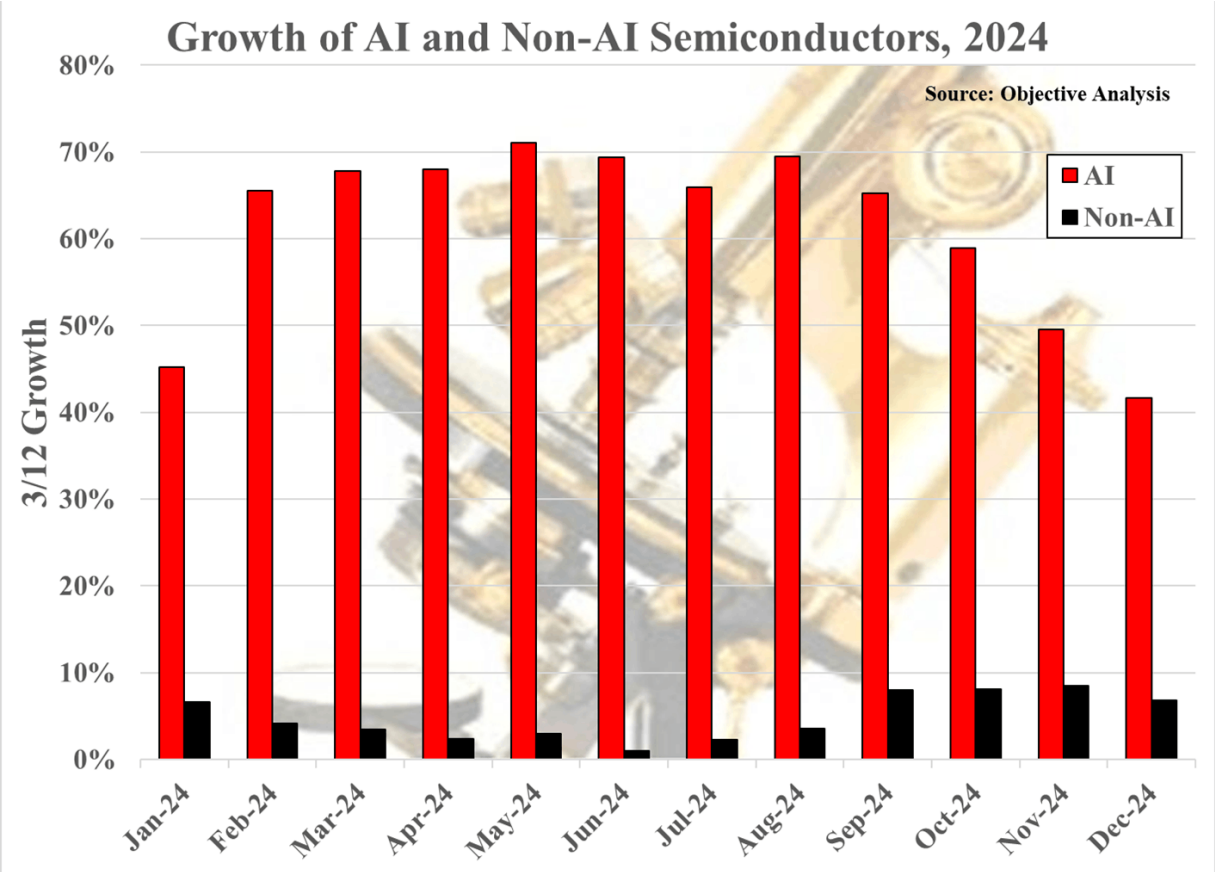
Figure 2 shows the incredible pace of this phenomenon, in which all major hyperscalers are participating. The handful appearing in this chart are those whose financials are easy to access, but others would add to this mix, too. The chart underscores such spending isn't the result of one company's quirks; rather, it's something that's being indulged in by the business at large.

Hyperscaler CapEx Surges, But This Has Troubling Implications

What's been troubling Objective Analysis for quite a while concerns hyperscalers purchasing new plants and equipment faster than their customers' AI spending has grown. How do you measure this? For that, we look at the capital spending above as a share of these companies' growing revenues.

If we add together all of these companies' revenues and divide it into the sum of their capital spending, we get their CapEx as a share of rev-

CHAPTER 4: What Happens When Hyperscalers Take Their Foot Off the Accelerator?

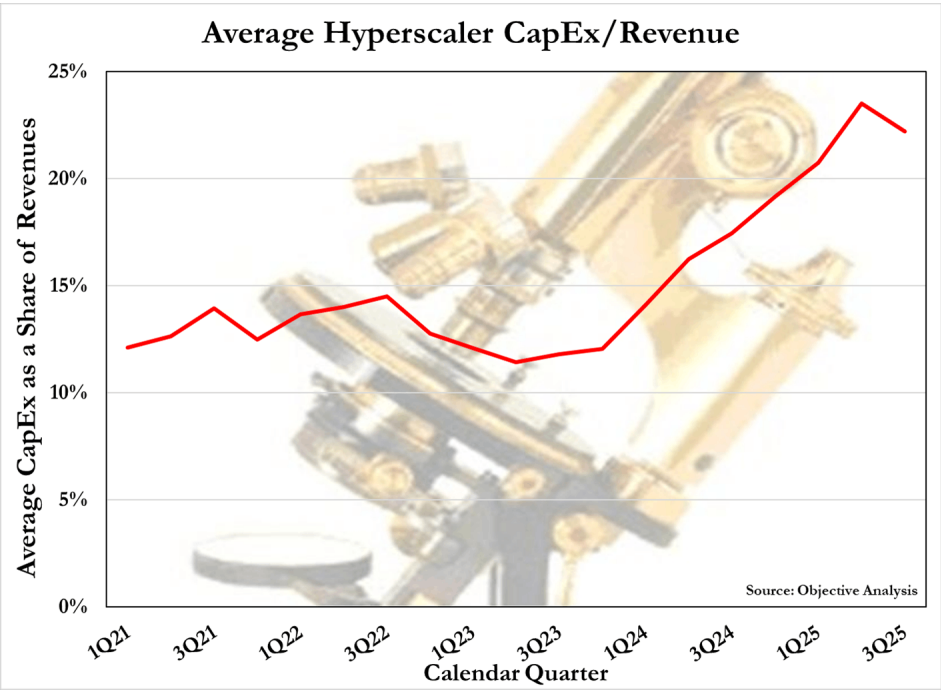


1. The difference between the growth of AI and non-AI semiconductors in 2024 is dramatically evident. Objective Analysis

enues. Corporate management runs companies according to their "Model," and that model specifically states what percent of revenues will go to CapEx, how much to research, etc. to bottom out at the profit they hope to deliver.

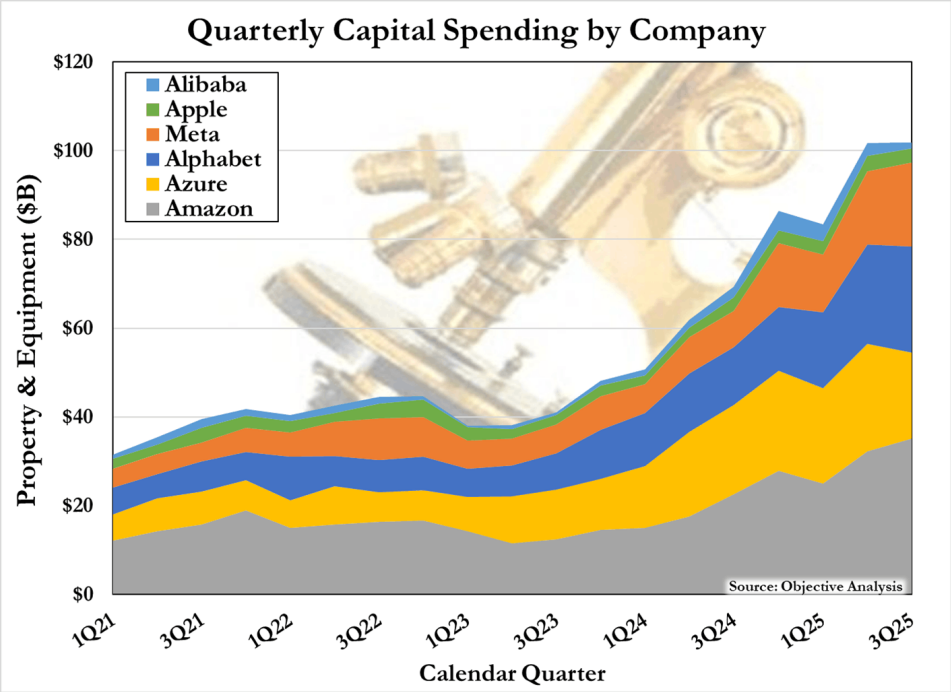
Investors pay a lot of attention to this, and voice their objections to management if profits suffer because some or one of these segments gets an unfair share of revenues. The six companies in the chart have spent, on average, between 12%-15% of their income on CapEx — at least until AI capital spending took off in 2024. Then the percentage suddenly almost doubled to 25% **(Fig. 3)**.

Although we haven't gone deeply into this area, Objective Analysis

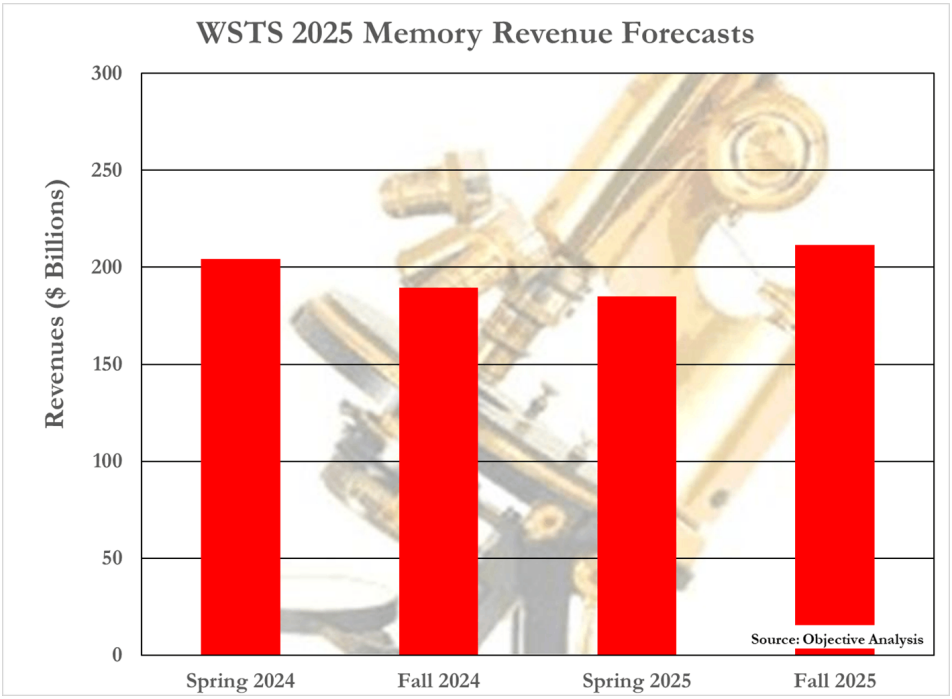


2. This chart shows the incredible pace of hyperscaler capital spending, with participation from all major hyperscalers. Objective Analysis

CHAPTER 4: What Happens When Hyperscalers Take Their Foot Off the Accelerator?



3. These six companies spent, on average, between 12%-15% of their income on CapEx — at least until AI capital spending took off in 2024. Then the percentage



4. The chart shows the 2025 revenue projection from the WSTS Spring 2024, Fall (Autumn) 2024, Spring 2025, and Fall 2025 forecasts. You can see that the forecasts go down and then back up again. Objective Analysis

suspects that management has convinced investors that they should take lower profits today for the promise of enormous benefits in the future. We haven't yet added SG&A data to our model, but we believe that it's being managed down to partly offset the capital spending increase.

Electronic Design readers are painfully aware of the fact that these companies have been laying programmers off, sometimes explaining that they're giving these tasks to AI. I don't know about you, but I've had numerous experiences with AI generating some pretty screwy results, so it doesn't surprise me at all that the latest couple of updates to Microsoft Office introduced very visible new bugs into the code.

The AI Quotient

But, back to the topic at hand, why does this matter to the chip market? It's because a sizable portion of this CapEx is being spent on AI semiconductors, which essentially means GPUs and HBM.

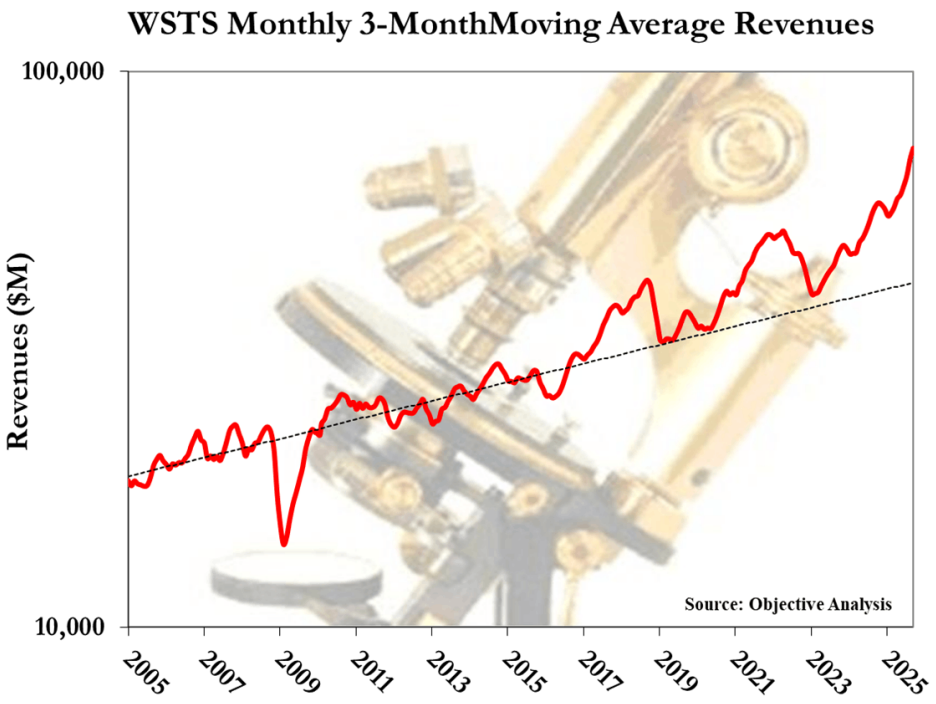
The companies participating in this have doubts just as we do at Objective Analysis. An easy way to tell is to look at the last four WSTS memory forecasts for 2025. The chart in **Figure 4** below shows the 2025 revenue projection from the WSTS Spring 2024, Fall (Autumn) 2024, Spring 2025, and Fall 2025 forecasts. You can see that the forecasts go down and then back up again. This gives us some important insights, since it's the leading memory companies' view of the future. Here's

why.

WSTS produces its forecast by consensus: The companies who ship the product compare their forecasts and decide how to incorporate each company's best ideas into the committee's forecast. In the case of the WSTS memory forecast, all memory makers who participate in this exercise have had a chance to air their views and agreed to publish this number.

If you look again at the CapEx numbers in *Figures 2 and 3*, you'll notice that the hyperscalers seem to be having second thoughts as well. Third-quarter capital spending, on the far right side of the chart, was roughly equal to second-quarter spending. It's largely due to a sizable decrease from Microsoft, which seems to be more cautious than other very aggressive spenders. (Some companies follow their own best thinking, little influenced by the crowd, with Apple being a prime example.)

We now have two points: AI CapEx is driving companies to make risky investments, investments that their investors may start to complain about, and the hyperscalers, their suppliers, and some analysts are beginning to question whether this can or should continue.



5. Semiconductor monthly revenues, on a log scale, are compared to the 3.9% average annual growth trend that these revenues have followed since 1996. The left side of the chart shows how closely these revenues generally followed the trend, with the exception of 2008-9, during the global economic collapse.

Objective Analysis

What if the AI Semiconductor Investments Don't Continue?

Figure 5 shows semiconductor monthly revenues, on a log scale (we use a log scale because this format shows constant growth as a straight line; in a standard linear format, it would take on the shape of a hockey stick), compared to the 3.9% average annual growth trend that these revenues have slavishly followed since 1996, or almost 30 years!

The left side of the chart shows how closely these revenues generally have followed the trend, with the noteworthy exception of 2008-9, during the global economic collapse. This caused a demand lapse, and such lapses were typically unusual enough that I would recite them and tell my audience that they occurred only every 15 years.

Then, in rapid succession, we had three demand cycles: a trade war scare that caused inventory builds and the 2018 cycle; then the COVID-19 pandemic, which resulted in another big cycle in 2022; and now we have AI driving another big cycle.

It doesn't take much imagination to figure out that this cycle is likely to end with the market returning to the trendline as it did not only in the "up" cycles of 2018 and 2021, but also in the 2008 downturn. It's something that the market has done repeatedly since 1996.

So, about that downturn: It's not a question of "if," rather it's a question of "when?"

And what will occur as a result?

As a memory analyst I can tell you: When DRAM shifts from a shortage to an oversupply, prices collapse to cost in about two quarters, which in many past cases have been a 60% drop. That's 60% in two quarters! It's an unsustainable change.

The same mechanism impacts NAND flash as well. Economists call it "The Commodity Cycle." Despite their leading-edge high-tech nature, DRAM and NAND flash are commodities. There's an explanation of this on "The Memory Guy" blog.

Rough Seas Ahead in 2026?

Is 2026 the year of the downturn? It's hard to tell. NVIDIA has done a wonderful job of lining up other customers to take up the slack if the hyperscalers decide to stop spending for a quarter or two. The most likely next one is various governments who have lots of money and are worried about "Data Sovereignty." They don't want some other country's data center to be supporting their AI needs, hosting all of their national secrets.

But if hyperscalers decide to slow their AI spending in 2026, it's highly likely that the memory market will collapse, and that GPU shipments will undergo a smaller-but-still-drastic decline.

Objective Analysis' outlook for 2026 is negative, but with the understanding that we may be very wrong. We simply find it difficult to expect hyperscalers to keep up their current spending level. Instead, we believe that investor pressure will drive them to pause and wait for demand to catch up with supply. This will result in canceled memory and GPU purchases, which will hurt GPUs and ravage the DRAM and NAND markets.

The timing is hard to predict, but the outcome is unavoidable.

view this article online

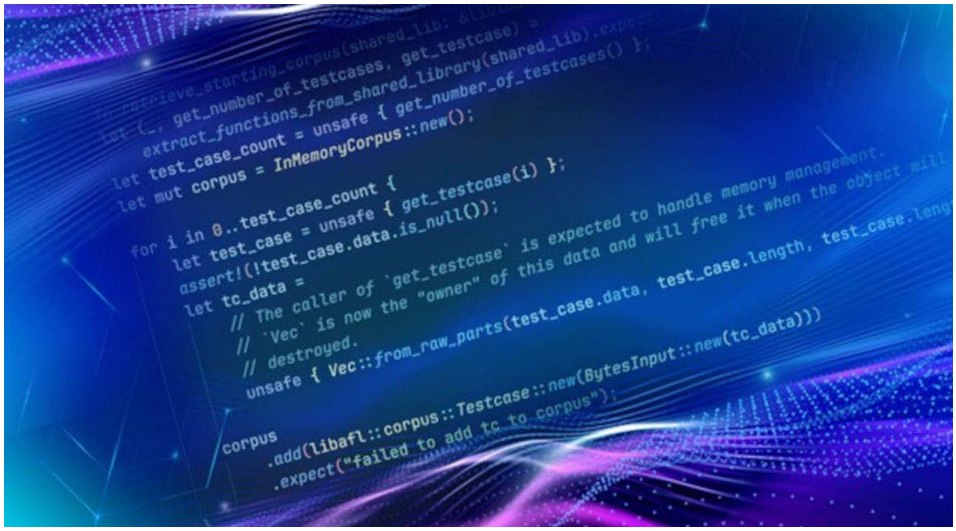
BACK TO TABLE OF CONTENTS

CHAPTER 5

Rust in Safety-Critical Systems: Predictions for 2026

As safety-critical industries move beyond experimentation and toward production-grade adoption of Rust, 2026 marks a pivotal transition from promise to practical assurance.

TONY AIELLO, Product Manager, *AdaCore*



AdaCore

Production-grade Rust is here and it looks to meet the requirement of safety-critical industries. While Rust's memory-safety guarantees are compelling, certification demands far more than safe code. It requires predictable execution, qualified toolchains, supply-chain trust, and auditable evidence. This article explores how Rust is evolving to meet these expectations, where early deployments are taking shape, and what organizations must do to translate innovation into a certifiable, long-term engineering advantage.

Regulators, primes, and Tier-1 suppliers increasingly expect memory-unsafe defects to be engineered out of critical paths. Rust's ownership and type system eliminate whole classes of vulnerabilities at compile time, but certification demands more than an absence of undefined behavior.

It requires disciplined processes, predictable execution, transparent supply chains, assured long-term support, and evidence that can be presented to an auditor. 2026 is the year when we stop asking whether Rust belongs in safety-critical software and start asking where it moves the assurance needle fastest.

State of Play

Pilots are real. Multiple organizations have shipped internal prototypes or non-critical modules in Rust. The common pattern: contain risk at system edges — parsers, communications, crypto, and device interfaces — before moving further into control logic.

Tooling is close. Static analysis, fuzzing, coverage, and timing analysis for Rust continue to mature, and formal toolchain qualification has started to enter vendor roadmaps.

Rust Predictions for 2026

The momentum behind Rust in the safety-critical domain is expected to continue growing through 2026, with adoption expanding into production-level subsystems. The early enthusiasm surrounding Rust's safety guarantees will give way to a more mature phase in which organizations balance innovation with certification constraints. As this shift occurs, sev-

eral trends are already becoming clear.

The first is the emergence of a practical, widely accepted subset of Rust that's safety-critical. While Rust's full feature set continues to evolve at high velocity, certifiable software developers will converge on a stable, conservative subset focused on predictable compilation, analyzable semantics, deterministic execution, and controlled use of advanced language features.

Much like MISRA C or SPARK subsets, this "Rust for Safety" profile will become the de facto baseline for certification, not imposed by any single authority but shaped organically through industry experience, shared guidelines, and tool support. As part of this consolidation, organizations will increasingly rely on long-term-supported toolchain versions, with vendors offering critical fixes on frozen branches rather than forcing updates to fast-moving upstream releases.

At the same time, mixed-language architectures will continue to be the standard practice. Most safety-critical teams will continue relying on large C, C++, or Ada codebases while selectively introducing Rust for components that benefit from its strong safety properties. This will drive growing demand for robust mixed-language tooling, particularly in areas such as cross-language debugging, static analysis, and coverage measurement.

Successful 2026 projects will be those that introduce Rust incrementally rather than attempting full rewrites.

A third trend is that toolchain qualification and library certification efforts will continue to solidify. By 2026, vendors will provide stabilized Rust toolchains suitable for qualification under DO-178C, ISO 26262, IEC 61508, and EN 50716. Early qualification kits will appear, covering compiler configurations, build-system workflows, package-manager restrictions, coding-standard enforcement, and coverage-analysis tools. In parallel, certifiable subsets of Rust run-time libraries will begin to emerge.

While these offerings will not yet be exhaustive, they will be enough for early adopters to launch real certification programs with a credible and auditable toolchain story. And they'll be backed by versioned, long-term-supported compiler branches with trackable defect histories and the ability to receive targeted fixes.

Ultimately, 2026 is anticipated to witness a significant rise in expectations regarding end-to-end supply chain trust. Safety-critical organizations will require signed, reproducible builds of Rust toolchains; verifiable SBOMs for all build components; and continuous vulnerability tracking across the Rust ecosystem.

This will move the industry beyond generic open-source trust toward a fully auditable supply chain that includes the compiler, its dependencies, the run-time libraries, and any crates allowed in restricted environments. Rust's package ecosystem will adapt with more crates offering provenance guarantees and long-term support branches.

Taken together, these developments suggest that 2026 will be the year Rust transitions from exploratory adoption to structured, certification-aware deployment. The pace will remain cautious, governed by engineering realities, regulatory expectations, and certification constraints. However, the foundations will be firmly in place for Rust to become a credible option for new safety-critical programs.

Rust Sector Implications and Examples

Aerospace & Defense

Rust's entry point in aerospace and defense will be lower-criticality systems, most commonly DAL C and DAL D modules. These levels offer a practical environment in which teams can accumulate certification evidence, validate toolchain usage, and qualify Rust workflows, laying the foundation for eventual use in higher-criticality DAL A/B functions.

Automotive

Automotive adoption is likely to advance faster than in aerospace and defense. The sector's entrenched use of C and C++, and the well-known memory-safety problems associated with them, creates strong pressure for safer alternatives, particularly as vehicle software grows in complexity and exposure. Rust's built-in guarantees directly address these weaknesses, making it an appealing candidate for early use in ASIL B/C components and for security-sensitive subsystems.

Conclusion

By 2026, Rust will have moved decisively from exploratory experimentation into early structured adoption within safety-critical industries. Its strong memory-safety guarantees and active ecosystem continue to attract significant interest. The next phase of adoption, though, will be shaped less by enthusiasm and more by the practical realities of certification, long-term support, and system integration.

Safety-critical programs will begin deploying Rust in narrowly scoped, lower-criticality components, DAL C/D in aerospace and defense, and ASIL B/C in automotive, where certification hurdles are manageable. Also, Rust's safety benefits can be realized without endangering program schedules. These early deployments will serve as proving grounds, generating evidence, experience, and confidence needed for future expansion into higher-criticality domains.

Across industries, mixed-language architectures will remain the norm. Rust will complement, not replace, existing C, C++, and Ada codebases, initially appearing in modules where memory safety provides a clear return on investment.

Meanwhile, the broader ecosystem will mature. A pragmatic, certifiable Rust subset will begin to take shape; qualification kits for stabilized toolchains will emerge; and early certifiable run-time subsets will be introduced. End-to-end supply-chain trust; signed, reproducible toolchains; verifiable SBOMs; and auditable dependencies will evolve from a "nice to have" to a procurement requirement.

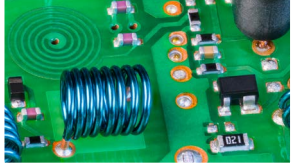
All of these developments indicate a cautious yet steady trajectory. Rust will not displace established languages overnight, nor will it achieve full certification readiness across all standards within a single year.

However, the foundations built throughout 2026 will position Rust as a serious and increasingly credible option for new safety-critical programs entering concept and early design phases. In this sense, 2026 will mark Rust's transition from promising newcomer to an emerging member of the safety-critical software ecosystem.

view this article online

BACK TO TABLE OF CONTENTS




☞ ElectronicDesign

Electronic Design has a diverse array of content for your perusal. To explore what our site offers, check out these links:

☞ Special Reports

☞ Top Stories of the Week

☞ Inside Electronics

☞ TechXchanges

☞ Newsletters

☞ Blogs

☞ Become an Electronic Design Member