

Rationed, Not Priced: What CoWoS Reveals About Bottleneck Power

Testing the constraint migration pattern at advanced packaging

When something becomes scarce, most executives look for the price increase. The more important signal may be who gets access, and who doesn't.

This is Article 3 of The Hidden Layer Series. Article 1, "The Map Nobody Has Drawn," laid out the AI Value Map. Article 2, "The Accidental Kingmaker," showed how bottleneck power migrates through the AI stack, using HBM as the case, and named CoWoS as the next test. This is that test.

CORE INSIGHTS

- CoWoS shows the first two stages of the HBM pattern: a hard constraint, and a supplier with real leverage.
- We underestimated it. Article 2 rated CoWoS Elevated. The evidence puts it in the High band, alongside HBM.
- The surprise: the leverage shows up as control over allocation, not yet as outsized margin. TSMC is rationing more visibly than it is maximizing price.
- Alternatives are being built. Customers have not yet moved meaningful volume to them.
- Advantage goes to companies that spot the next constraint before the current one resolves.

THE BUSINESS LESSONS, IN PLAIN LANGUAGE

- **Money doesn't buy your place in line.** When something is truly scarce, the supplier decides who gets it. Your budget matters less than your position in the queue.
- **Power doesn't always show up as a higher price.** The bigger risk isn't paying more. It's getting less, or getting it later.
- **An alternative existing isn't the same as one you can use.** Count on relief only when customers have actually moved volume to it.
- **Bottlenecks move; they rarely disappear.** Solving today's constraint usually exposes the next one. Plan for that one.

The Promise We Made

In Article 2, we made a bet in public. We said the HBM bottleneck would not disappear. It would migrate, and the next stop was advanced packaging.

Specifically, CoWoS: TSMC's Chip-on-Wafer-on-Substrate, the process that bonds an AI accelerator to its memory. For many of today's highest-end AI accelerators, CoWoS remains the dominant route to high-volume shipment.

We called CoWoS a candidate, not a conclusion. And we said Article 3 would check.

In July 2026, TSMC's CEO answered part of the question for us. On the company's second-quarter earnings call, C.C. Wei said packaging capacity is "so tight that now it's limiting my customers' growth."

That confirms a constraint. What it reveals about power turned out to be more interesting than we expected.

The Test



The Hidden Layer: The AI Value Map. 13 layers, each with its purpose and its chokepoint.

Article 2 asked three questions of every bottleneck. This article sharpens them.

THE M37 LENS: Three questions, not one

Constraint. Can the market get enough supply?

Bottleneck power. Does the constrained supplier control something customers cannot easily substitute?

Value capture. Does that power turn into price, margin, allocation control or strategic advantage?

Article 2 summarized this as Bottleneck Power = Constraint × Market Structure × Customer Urgency. Treat that as a heuristic, a way of thinking, not an equation.

Here is where the evidence through September 2026 stands.

WHAT WE KNOW, WHAT'S EMERGING, WHAT WE'RE TESTING

Known: CoWoS is a binding capacity constraint, and TSMC holds the leverage.

Emerging: That leverage is being used to allocate and ration, with price rising but margin capture not yet proven. Alternative packaging is being built.

Not yet demonstrated: Material customer volume moving away from CoWoS.

Three Tests of the Pattern

① The Constraint Held, Even as Capacity Roughly Quadrupled

TSMC has built faster than almost anyone expected. Industry estimates compiled by [TrendForce](#) put CoWoS at about 35,000 wafers a month in late 2024. End-2026 estimates now range from about 90,000 (TrendForce's earlier projection) to 140,000 (investors cited by [Economic Daily News](#)). TSMC itself has guided to more than 80% annual growth from 2022 to 2027.

It still isn't enough. The same investors estimated the supply gap at about 20% in mid-2026, narrowing to roughly 10% by year-end. Analysts now see TSMC [doubling again to about 260,000 wafers a month by end-2028](#). That is an analyst estimate, not TSMC guidance.

Why does a build-out this large still fall short? The chips got bigger. Each new generation of AI package is larger, so every wafer yields fewer of them. Capacity grows in wafers. Demand grows in silicon area.

② The Leverage Is Real. It Isn't Showing Up as Margin.

This is where the evidence surprised us.

Price has risen. A CoWoS wafer now reportedly sells for [about \\$10,000, on par with a 7nm logic wafer](#). Allocation is concentrated: Morgan Stanley estimated NVIDIA had secured [roughly 60% of 2026 capacity](#).

But the same TrendForce report says TSMC's packaging margins still trail its corporate average. And when asked about Intel's rival packaging, TSMC's CEO said: "We hope they will be successful."

A dominant supplier that welcomes overflow capacity is not behaving like one squeezing every dollar from scarcity. TSMC is using its leverage to manage allocation, while its front-end wafer business remains the larger economic engine.

“Bottleneck power doesn't always show up as price. Sometimes it shows up as the power to decide who waits.”

Our original model assumed scarcity leads to leverage, then price, then profit. CoWoS suggests a second path: scarcity leads to leverage, then allocation control, then strategic rationing.

Allocation power is the ability to decide who receives scarce capacity, when they receive it and on what terms, even when the supplier is not maximizing price. It applies well beyond chips: to cloud capacity, energy, talent and channel capacity.

③ Alternatives Are Being Built. Volume Hasn't Moved.

Stage 3 of the pattern is migration: competition and capital attack the bottleneck. Building an alternative is not the same as customers using it.

What we're seeing	What it means
Intel EMIB-T design wins, including a reported MediaTek dual strategy for a future AI ASIC	Alternative capacity emerging; analysts estimate 15,000–20,000 CoWoS-equivalent wafers a month in 2027
OSAT partners adding an estimated 50,000–60,000 wafers a month	Additional packaging capacity; much of it positioned as overflow or supplemental capacity
TSMC's panel-level packaging (CoPoS)	Longer-term alternative; volume production not expected until around 2029
High-volume displacement of CoWoS for flagship accelerators	Not yet demonstrated

The lesson extends beyond chips: don't count competitors; count qualified alternatives with actual volume.

The Scorecard: What Article 2 Got Right and Wrong

We promised to mark our own work, including the misses.

Article 2 hypothesis	Evidence through Sept 2026	Status
CoWoS shows the structural conditions for a bottleneck	Dominant supplier, sold-out capacity, multi-year build-out	✓ Confirmed
Amplifiers are weakening (urgency and capital cooling)	Hyperscaler 2026 capex up about 77% (FT tally, Q2 guidance); TSMC says packaging limits customer growth	✗ Not supported
Pricing power becomes visible	Price near 7nm levels; packaging margins still below TSMC's average	⚠ Partly confirmed
Competitors begin qualifying	EMIB-T design wins, OSAT capacity, TSMC's own bridge program	✓ Confirmed, early

Article 2 hypothesis	Evidence through Sept 2026	Status
Supply gap narrows from ~20% to ~10% by end-2026	Forecasts on track; year-end data not yet in	 Open

“The structural conditions were right. The amplifiers were wrong.”

We assumed AI spending would cool and customers would push back. They didn't. That is why we underestimated CoWoS.

Bottleneck Power Score: CoWoS, Re-Rated

M37 Directional Research Index. From this article on, M37 reports bands rather than decimals: High (8-10), Elevated (7-8), Moderate (5-7), Low (below 5). Article 2's composite scores were judgment-weighted; bands make comparisons across articles consistent. The index is directional, not statistical: it compares the strength of structural conditions across bottlenecks, and does not predict financial outcomes.

Dimension	Article 2 (forecast)	Article 3 (Sept 2026 evidence)	What changed
Supplier Concentration	9	9	TSMC still dominant
Qualification Time	8	8	Unchanged
Replication Cost	9	9	Unchanged
Switching Cost	6	6	EMIB is an option for some ASICs, not flagship GPUs
Customer Expected-Return Urgency	5	9	TSMC's CEO says capacity limits customer growth; customers are supply-limited, not price-sensitive
Capital Availability	6	9	Hyperscaler 2026 capex up about 77% year over year
Band	Elevated	High	Same band as HBM 2024-25

Why This Pattern Keeps Repeating

Two tests, same early shape: HBM and CoWoS both moved from constraint to bottleneck power. HBM has since moved further toward migration; CoWoS is building the conditions for it.

That makes constraint migration a working pattern worth testing, not a law. Constraints rarely disappear. More often, they move. And patterns let you look ahead.

That is what the AI Value Map is for. It is not a map of where value sits. It is a map of where constraints, and eventually bargaining power, can migrate.

Where does the scarcity go when packaging catches up? TSMC can double CoWoS by 2028. The question is whether the buildings those chips go into can be energized as fast. Article 4 tests whether the next binding constraint is power, or whether another layer captures the scarcity first.

What To Do Next

Start with three questions:

- 1. Where are you in line?** Ask your cloud and server vendors whether your AI capacity depends on constrained packaging, and where you sit in the allocation queue. Roadmaps don't tell you that.
- 2. Do you have a second path?** If you design custom silicon, an EMIB-T or outsourced packaging route may be slower to qualify but cheaper than waiting. If you buy capacity, know which alternative providers you could switch to.
- 3. Are you planning for the next constraint or the last one?** If packaging eases in 2027, what stops you then?

Then act on the answers:

Reader	What to do
IT and infrastructure leaders	Ask vendors for committed delivery dates and allocation status, not roadmaps. Build schedule buffers into AI projects that depend on new accelerators.
Boards and CFOs	Add one question to every AI investment approval: "What is our binding constraint, and do we have access to it?"
Procurement and supply chain	Map single-source dependencies. Start qualifying second sources early; it can take a year or more. Time renegotiations to when alternatives qualify.
Technology partners	Set customer delivery expectations early, and offer interim options, such as cloud capacity, while hardware is allocated.
Everyone	Track the four signals in the prediction table below each quarter. Each one shows the next shift before it reaches earnings.

The Prediction

Same rules as Article 2: stated now, checkable by anyone, marked in public.

Read Article 2: The Accidental Kingmaker

Prediction (by end-2027)	Confirms	Falsifies	Status
TrendForce or equivalent supply-chain trackers report CoWoS lead times under 52 weeks	Published lead-time data below 52 weeks	Lead times stay at 52+ weeks	Open
At least one accelerator from NVIDIA, AMD, Google, Amazon or Broadcom ships more than 100,000 units on EMIB-T or another non-CoWoS 2.5D package	Reported volume shipments above that level	All flagship accelerators remain on CoWoS	Open
At least two of Microsoft, Alphabet, Amazon and Meta cite power or grid access, not packaging, as their primary capacity constraint on two consecutive earnings calls	Earnings-call language shifts to power	Packaging remains the named constraint	Open

Prediction (by end-2027)	Confirms	Falsifies	Status
TSMC's advanced packaging gross margin reaches or exceeds its corporate average	Margin parity or better reported	Margin stays below average	Open

If the first three hold, CoWoS will have reached Stage 3, and Article 4's power hypothesis gains support. If the fourth fails, our read that CoWoS bottleneck power would translate into stronger margin capture was wrong.

THE BOTTOM LINE

Scarcity creates power, but power doesn't always show up as price. Sometimes it shows up as access: who gets supply, and who waits.

Constraints rarely vanish; they move. The advantage goes to companies that spot the next one first.

Where is packaging stopping you? What did we score wrong? Show me your evidence. The map improves when you do.

How This Article Was Tested

Before publication, every number, causal claim and prediction in this article went through a claim-by-claim devil's-advocate review. Five claims were corrected as a result, and the central finding, that bottleneck power can show up as allocation rather than price, came out of that review.

Practitioners in advanced packaging, semiconductor supply chains and AI infrastructure: if you see where the evidence is thin, say so in the comments. Challenges that change the analysis will be credited by name in the next update.

ABOUT THE AUTHOR

Joe Batista advises boards, C-suite executives, and go-to-market teams on technology value creation, growth strategies, and the opportunities hidden beneath major technology shifts. He has spent more than 40 years in the technology industry, guiding companies on growth and technology decisions, spanning the minicomputer era at Digital Equipment Corporation (DEC) through client-server, cloud, virtualization, and now AI, with executive experience at Dell Technologies and Hewlett Packard Enterprise.

Through M37 Advisory, Joe helps leadership teams recognize recurring patterns, uncover second-order effects, find the opportunities others overlook, and identify the strategy worth betting on. His work is grounded in a simple belief: **What if the biggest obstacle to your growth isn't the competition, but the way you think?**

"Most executives have been bombarded with 'theories' about what might work. Based on our research, Joe's extensive playbook is rooted in creating alignment of interest and business outcomes that leaders should take notice of."

— Jim Dickie, Research Fellow & Sales Mastery Founder, CSO Insights

Author Contact

Email: Chief.Creatologist@M37-Advisory.com

Website: <https://www.m37-advisory.com>

Idea Vault: <https://www.m37-advisory.com/idea-vault>

NEXT IN THE SERIES

Article 4: The Transformer Bottleneck — We test whether the constraint moves down the stack to Layer 0, where power, transformers and grid interconnection decide what ships.

To subscribe to the series, [complete the form](#)

Sources

Company statements

- [TSMC Q2 2026 earnings call transcript](#) (Jul 2026)

CoWoS capacity and supply gap

- [TrendForce: TSMC ramps up CoWoS capacity across Taiwan](#) (Dec 2024)
- [TrendForce: CoWoS supply-demand gap narrowing to 10% by end-2026](#) (Jun 2026)
- [TrendForce: CoWoS to double by 2028; EMIB-T capacity estimates](#) (Sep 2026)

Pricing, margin and allocation

- [TrendForce: CoWoS wafer ASP nears 7nm](#) (Apr 2026)
- [Astute Group: Morgan Stanley on NVIDIA's CoWoS allocation](#)

Competitive packaging moves

- [TrendForce: MediaTek adopts dual EMIB and CoWoS strategy](#) (May 2026)
- [Tech Times: TSMC starts EMIB-like program](#) (Jul 2026)

Demand

- [AI Weekly: Financial Times tally of 2026 hyperscaler capex](#)

Keywords: AI infrastructure, CoWoS, advanced packaging, HBM, bottleneck power, value capture, constraint migration, AI Value Map, M37 Advisory