

Semiconductor Weekly News

August 24 – 30, 2026

Custom silicon showed up at every scale at Hot Chips — from 700W data-center ASICs to robotaxi front ends.

Miki Matsui, Co-Founder, CHIP-J

Sources: NVIDIA, Apple, Waymo, Hot Chips, SEC, The White House, Tom's Hardware, CNBC and others

Editor's Picks

\$96.2B

**NVIDIA Q2
FY2027 revenue**

Data Center reached \$89.0B, up 117% year over year. Q3 guidance is \$108.0B $\pm 2\%$ on the Vera Rubin ramp — and assumes no China data-center compute revenue at all. ([NVIDIA, Aug 26](#))

1.9x

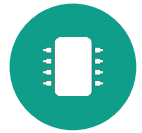
**OpenAI's first
custom ASIC**

OpenAI's first custom inference chip. On a public benchmark it reports 1.5x-1.9x more work per watt than Nvidia's 1,200W GB200 and 1,400W GB300. ([Open AI, Aug 25](#))

20x

**Waymo
compute, 8
years**

Waymo disclosed its custom 5nm sensor-fusion processor at Hot Chips, citing more than 200 million fully autonomous miles. ([Waymo, Aug 24](#))



AI Chips & GPUs

1 **NVIDIA reports \$96.2B for the quarter**

Revenue for the quarter ended July 26 rose 106% year over year to \$96.2B, with Data Center at \$89.0B. Gross margin held at 75.0% and diluted EPS was \$2.46. The release credits Vera Rubin, now in full production, rather than Blackwell. ([NVIDIA, Aug 26](#))

2 **NVIDIA Groq 3 LPX enters full production**

Announced from Hot Chips as "the interactive AI inference accelerator." Artificial Analysis measured 3,400 output tokens per second on Gemma 4 31B at a 100,000-token context. ([NVIDIA, Aug 24](#))

3 **OpenAI details the Jalapeño inference ASIC**

On SemiAnalysis's public InferenceX benchmark OpenAI reports 1.5x-1.9x more work per watt and 1.7x-3.6x lower latency than GB200 and GB300, rising to 2.1x-4.1x on highly interactive workloads. ([OpenAI, Aug 25](#))

4 **Cerebras maps the next wafer-scale generations**

CS-4 is the first system on the Nexus rack platform, with 6x the memory bandwidth, compute and IO of CS-3. Cerebras says CS-6 "will push the frontier of 3D integration by stacking DRAM at wafer-scale." ([Cerebras / Hot Chips, Aug 26](#))



Fab & Supply Chain

1 Micron: HBM's silicon penalty widens each generation

HBM consumes roughly three times the wafer area of DDR5 for the same capacity, and the position is "definitely not getting better." Micron sees a good path to 16 DRAM layers, with much unresolved beyond that. HBM4E moves the base die to TSMC. ([Hot Chips, reported Aug 26](#))

2 SK hynix pushes hybrid bonding out to HBM5

The company will skip hybrid bonding for HBM4E and keep MR-MUF through the Nvidia Rubin generation. Bonding 16 to 20 dies reliably remains unsolved; its slides put the stack-height ceiling at 775 microns, up from 720. ([Hot Chips, Aug 25](#))

3 Intel says 14A is beating its defect-density target curve

CFO David Zinsner said Intel has "not seen this performance since 22nm" — a claim about the rate of improvement two years before volume, not about absolute yield today. Risk production is set for H2 2027 and volume manufacturing for 2028. ([Deutsche Bank conference, Aug 26](#))



Consumer Chips

1 Apple ships its first 2nm silicon in the Mac mini

Apple describes M6 as its first 2 nm chip, with a 12-core CPU — two more cores than M5. Against the previous Mac mini with M4 it claims up to 4x faster AI and 40% faster CPU performance. From \$899, available September 22. ([Apple, Aug 25](#))

2 Samsung puts compute inside LPDDR5X

Presented as the world's first LPDDR-based processing-in-memory solution. Reported figures are 2.28x faster inference speed and 3.01x higher token throughput against standard LPDDR5X, at 614 GB/s versus 76.8 GB/s. ([Hot Chips, Aug 25](#))

3 Raymond James sees AMD passing Intel by 2027

Simon Leopold upgraded AMD to Strong Buy and raised his target to \$641 from \$565, arguing agentic AI is the principal new growth engine and modelling a CPU market of roughly \$201B by 2030. ([CNBC, Aug 25](#))



Hyperscalers, Data Centers & Energy

1 Hot Chips becomes a hyperscaler silicon showcase

Google presented an eighth-generation TPU family split into training and serving chips. Microsoft's MAIA 200 uses TSMC 3nm with over 140 billion transistors and 216GB of HBM3E in a 750W envelope. Meta introduced MTIA 400. ([Hot Chips, Aug 24-25](#))

2 NVIDIA pitches site-level power control for AI factories

Its DSX MaxLPS approach manages how an entire AI factory ramps power rather than optimising individual racks. As grid operators increasingly price interconnection on peak behaviour, a controllable ramp profile is becoming a siting asset. ([Hot Chips, Aug 25](#))

3 A \$250M contract shows the scale of West Texas build-out

Target Hospitality signed a multi-year lease and services agreement for a top-five hyperscaler's data-center development in the Pecos region, worth about \$250M of revenue through August 2030 against under \$15M of capital investment. ([SEC filing, Aug 26](#))

Autonomous Driving & Automotive Chips

1 Waymo discloses its sensor-fusion processor

Waymo took two slots on the opening day. The custom 5nm part is not the driving brain: it sits ahead of the main ML compute, denoising and running early networks on raw lidar, radar and camera streams. ([Hot Chips, Aug 24](#))

2 "Over 1,000 TOPS" is the system, not the chip

Waymo's blog wording was widely read as a per-chip figure and compared to Nvidia's DRIVE AGX Thor. The company confirmed it describes the full system. Vendor TOPS claims in this segment are not yet quoted on a common basis. ([TechCrunch, Aug 30](#))

3 Waymo begins laying the groundwork in Munich

The company says it is "laying the groundwork" for fully autonomous ride hailing in Munich, with commercial service targeted toward the end of 2027. Near-term work is manual driving and HD mapping. ([Waymo, Aug 25](#))

4 Gatik raises a \$200M Series D

Led by the Qatar Investment Authority and Koch Disruptive Technologies with ARK Invest participating. Its PepsiCo partnership, announced in June, covers 41 driverless trucks; company-wide contracted revenue exceeds \$600M. ([TechCrunch, Aug 25](#))



Chip Startups & Venture

\$40M

Quintessent — Series A

Led by Cycle Capital with Goldman Sachs and Ciena participating. The company integrates a quantum-dot DWDM comb laser onto silicon photonics for AI data-center optical interconnect. ([BusinessWire, Aug 24](#))

\$30M

Celera Semiconductor — Series B

Funded entirely by Maverick Silicon, its largest investor. Celera calls Nesto the industry's first library of digital twins of analog functions. ([PR Newswire, Aug 25](#))

\$24.5M

Agentrys — seed plus pre-seed

A \$19.1M seed led by Etna Labs follows a \$5.4M pre-seed led by MediaTek. The platform is pitched as full-flow agentic design automation, with the new money going to verification and physical design. ([Agentrys, Aug 26](#))

100 TB/s

d-Matrix shows Raptor

A TSMC 4nm compute die is bonded face-to-face at 36-micron pitch onto a custom DRAM die, giving 100 TB/s from 32GB per card. It was shown in a tutorial, not the main program. ([Hot Chips, reported Aug 26](#))



Industry Deals & Policy

NVIDIA reported in talks for Hugging Face

The Information reported agreement at \$12.9B; Business Insider reported unsigned talks above \$13B. Neither company has commented, and no primary source exists. ([CNBC, Aug 27](#))

US weighs far broader chip tariffs

Duties under discussion reach beyond chips to laptops, servers and gaming hardware, with duty-free volumes reportedly pegged to each firm's committed US production rather than

National emergency declared over grid equipment

Executive Order 14420 gives the Energy Secretary 120 days to write rules covering substation transformers, inverters, storage and breakers. It reaches equipment already installed and names no country. ([EO 14420, Aug 26](#))

Export controls now baked into the guidance

NVIDIA's third-quarter outlook assumes no China data-center compute revenue at all — controls absorbed as a permanent modelling assumption rather than a temporary drag. ([NVIDIA, Aug 26](#))



Sources

- [NVIDIA — Q2 FY2027 results \(Aug 26\)](#)
- [NVIDIA — Groq 3 LPX in full production \(Aug 24\)](#)
- [OpenAI — Jalapeño first benchmark results \(Aug 25\)](#)
- [Waymo — sensor-fusion processor disclosure \(Aug 24\)](#)
- [Hot Chips 2026 — conference program \(Aug 23-25\)](#)
- [Apple — Mac mini with M6 and M5 Pro \(Aug 25\)](#)
- [The White House — Executive Order 14420 \(Aug 26\)](#)
- [SEC EDGAR — Target Hospitality 8-K \(Aug 26\)](#)
- Secondary: Tom's Hardware, ServeTheHome, CNBC, TechCrunch. Fact-checked against primary sources on Aug 30, 2026.

Every source citation on these slides links to the underlying article.