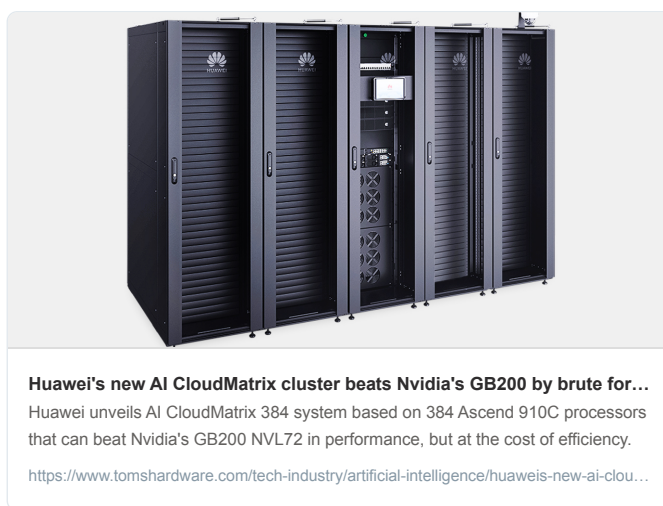




William Huo @wmhuo168

Apr 23, 2025 · 13 tweets · [wmhuo168/status/1914982988799943100](https://twitter.com/wmhuo168/status/1914982988799943100)

BREAKING: Huawei can't buy Nvidia's GB200 AI chips.
So it built something bigger.
Introducing CloudMatrix 384—a Chinese hyperscale cluster with 300+ PFLOPs of AI power.
It's not imitation.
It's escalation.
The West was not ready for this.



Huawei just deployed CloudMatrix 384—a datacenter rack system powered by its own AI chips, optical interconnects, and NVLink analogs.
You know the Nvidia GB200 NVL72?
China just answered that.
From scratch. 1/12

Each CloudMatrix 384 rack contains:

8x Ascend 910B chips per node

384 total chips

300+ PFLOPs AI compute

Built-in optical backplane
No InfiniBand. No Nvidia. No help.
All Huawei silicon. 2/12

How did they do this under sanctions?
Two reasons:

1. Chiplet design—910B is modular, scalable, and optimized for interconnects

2. Silicon-photonics—Huawei leapfrogged copper with in-house optical links 3/12
Western analysts laughed when China said it would go all-in on chiplets and optical IO.
Now Huawei's CloudMatrix scales faster than Nvidia's DGX or Grace Hopper racks.
This isn't catch-up. It's an end-run. 4/12

Huawei's 910B isn't some poor man's H100.
It supports FP16 and BF16 natively,
trains LLMs efficiently,
and already powers models from SenseTime, iFlyTek, and Baidu. 5/12

Each rack uses a high-speed optical backplane.
Not copper.
That means:

Lower latency

Easier scaling

Better thermal performance
And no Western equivalent at this density. 6/12

This isn't vaporware.
Huawei is mass-deploying CloudMatrix to over 40 datacenters.
That's GPT-4 scale compute.
Domestic.
No Nvidia.
No ASML.
No mercy. 7/12

And the U.S. can't stop it.
Huawei owns the entire vertical stack:

Chips

Boards

Interconnects

Racks

Framework (MindSpore)
No foreign licenses required. 8/12

The CHIPS Act was supposed to halt China's AI rise.
Turns out, it accelerated it.
No H100s?

Huawei said:

We'll just build our own hyperscaler. 9/12

CloudMatrix 384 is Huawei's answer to Nvidia's GB200 NVL72.

Except it's built in China, with Chinese fabs, optics, firmware, and funding.

And it's already shipping. 10/12

Next up:

Silicon Secrets Ep. 3 – "SMIC's Secret 5nm: DUV and the Death of Moore's Law"

China made 5nm chips without EUV,

using brute-force DUV + SAQP.

Physics? Optional when you scale like this. 11/12

Follow for the rest of Silicon Secrets.

This isn't a chip war.

It's the collapse of an illusion.

And the West is two tech cycles behind. 12/12

• • •