



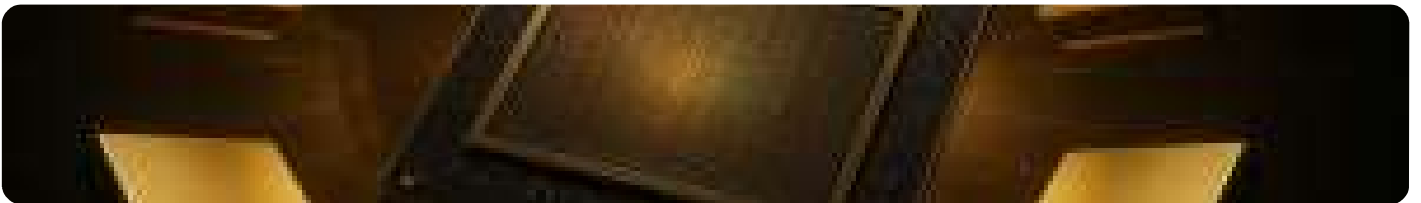
May 18, 2026

# Broadcom Pumafish Canceled, TPU v10 Icefish Pulled In, Nvidia Update, ByteDance ASIC

// SemiAnalysis Accelerator Model Update – May 18, 2026

3 minutes

By Myron Xie, Ivan Chiam, Louis Lu, Michael Chen and Ray Wang



[Download Model](#)

## Broadcom TPU 10 Icefish Pulled In, Broadcom TPU 9 Pumafish Canceled, BRoadcomTPU 8i+ Sunfish, MediaTek TPU 10

According to industry chatter Google and Broadcom's TPU 9 project 'Pumafish' / 'Bladerunner' has been cancelled so that Broadcom can pull in the TPU 10 'Icefish' schedule which will come directly after TPU 8i 'Sunfish.' Icefish is expected to tapeout mid 2027, which would mean Icefish would begin shipping mid 2028 if it's on schedule, which is only slightly later than Pumafish late 2027/early 2028. We believe the reason for this is to focus Broadcom and Google's design efforts on shipping Icefish earlier rather than have Pumafish in between. TPU 8i's Sunfish lifecycle will be extended until Icefish ships and there could be a TPU 8i+ upgrade which could be a dual Sunfish MCM in a single substrate (similar to Rubin Ultra) as a mid-cycle upgrade before Icefish.

MediaTek's TPU v9 'Humufish' / 'Helios' will be the only TPU 9 chip, which is expected to tape out in 2H 2026. MediaTek will continue to be involved in the follow on chip with Helios Next, its version of TPU 10 which we discuss below. Helios Next is expected to tape out in 2H 2027 (after Icefish). The layout is very similar to Broadcom's Pumafish / Bladerunner, which suggests that Helios Next is a replacement for Pumafish, even though Helios Next should ship after Icefish based on tape out schedules. Therefore, the roadmap timeline should be Humufish (MediaTek), Icefish (Broadcom), Helios Next (MediaTek).

Overall, we interpret this as Google wanting to accelerate TPU iteration and have Broadcom and MediaTek ship new generations subsequently after each other rather than have 2 different tracks launch simultaneously as is the case for v8. Between Broadcom and MediaTek, we expect Broadcom will continue to handle designs that are more aggressive architecturally. Overall, things appear to be trending very well for the Google and MediaTek relationship, with the next major milestone coming soon in terms of seeing whether the TPU 8t rework was successful to move on to ramping it for mass production.

Google TPU Roadmap			
Chip	Design Partner	Tape Out	Mass Production Timing
TPU 8i (Sunfish / Hellcat)	Broadcom	Already taped out	Currently shipping (lifecycle extended until Icefish)
TPU 9 (Pumafish / Bladerunner)	Broadcom	Canceled	Canceled
TPU 10 (Icefish)	Broadcom	Mid 2027 (expected)	Mid 2028
TPU 8t (Zebrafish)	MediaTek	Already taped out	2H 2026
TPU 9 (Humufish / Helios)	MediaTek	2H 2026 (expected)	2H 2027
TPU 10 (Helios Next)	MediaTek	2H 2027 (expected)	2H 2028

Source: SemiAnalysis

TPU 10 (Helios Next) is expected to adopt hybrid bonding with 8 base dies on N2P and 16 top dies bonded 2-on-1 (two top dies per base die), and 6 I/O dies. HBM remains undecided between HBM4E and HBM5, with the HBM5 option potentially being a custom implementation. Both EMIB and CoWoS are possible as the 2.5D integration scheme but we lean towards CoWoS given CoWoS should be ready for >9x reticle CoWoS by that timeline as well as MediaTek hiring Douglas Yu, who was a key member of TSMC’s advanced packaging pathfinding.

### TPU v10 (Helios Next)



Source: SemiAnalysis

# Nvidia – H200 Adjustments

The H200 stalemate continues. The US government has approved H200 export licenses to 10 Chinese companies to purchase up to 75k H200s each per approved company but the Chinese government has not yet authorized the purchases. For a brief moment, there were official CCP press releases saying H200 purchases were approved, but they were promptly taken down. Accordingly, we cut H200 to 50k units shipped (outside of China) in Q1 2026 and we also add 15k GB300 units in the same quarter. In Q2, we push out 200k H200 shipments (100k each to Q3 and Q4), this is again based on the assumption that Chinese approvals come through. We also increase our GB300 shipment estimates by 75k in Q2 2026. Ex-H200, our estimate for Q2 FY2027 Nvidia whole company revenue is \$95bn.

	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26	1Q27	2Q27	3Q27	4Q27
<b>DEMAND</b>												
<b>18 May 2026 update</b>												
B100/200/300 Units (CoWoS-L)	696,236	771,968	1,004,500	1,253,064	1,540,440	1,725,040	1,348,696	994,504	332,520	262,480	112,504	112,504
B200A/300A Units	0	0	0	0	0	0	0	0	0	0	0	0
B30A Units	0	0	0	0	0	0	0	0	0	0	0	0
H100/H200 Units	219,568	227,271	111,400	30,000	50,000	100,000	400,000	200,000	0	0	0	0
H20 Units	454,000	59,528	4,200	832	0	0	0	0	0	0	0	0
RTX Pro 6000 Blackwell Server Edition Units	0	0	200,000	250,000	150,000	0	0	0	0	0	0	0
Rubin Units	0	0	0	0	0	60,608	413,976	830,000	1,603,280	1,657,880	1,297,640	469,304
Rubin Ultra Units	0	0	0	0	0	0	0	0	8,352	256,176	698,944	0
Vera CPU Units	0	0	0	0	0	25,704	175,572	348,228	733,500	774,052	719,324	543,259
CPX Units	0	0	0	0	0	0	0	0	61,357	57,000	102,164	0
Nvidia Groq LPU units	0	0	0	0	0	0	120,000	230,000	300,000	300,000	300,000	300,000
% change vs previous												
B100/200/300 Units (CoWoS-L)	0%	0%	0%	0%	1%	5%	0%	0%	0%	0%	0%	0%
B200A/300A Units	-	-	-	-	-	-	-	-	-	-	-	-
B30A Units	-	-	-	-	-	-	-	-	-	-	-	-
H100/H200 Units	0%	0%	0%	0%	-50%	-67%	33%	100%	-	-	-	-
H20 Units	0%	0%	0%	0%	-	-	-	-	-	-	-	-
RTX Pro 6000 Blackwell Server Edition Units	-	-	0%	0%	0%	-	-	-	-	-	-	-
Rubin Units	-	-	-	-	-	0%	0%	0%	0%	0%	0%	0%
Rubin Ultra Units	-	-	-	-	-	-	-	-	0%	0%	0%	0%
Vera CPU Units	-	-	-	-	-	0%	0%	0%	0%	0%	0%	0%
CPX Units	-	-	-	-	-	-	-	-	0%	0%	0%	0%
Nvidia Groq LPU units	-	-	-	-	-	-	0%	0%	0%	0%	0%	0%
B100/200/300 Revenue	23,732	27,802	38,942	48,362	62,266	70,577	55,524	41,058	14,566	11,474	4,725	4,725
B200A/300A Revenue	0	0	0	0	0	0	0	0	0	0	0	0
B30A Revenue	0	0	0	0	0	0	0	0	0	0	0	0
H100/H200 Revenue	5,200	4,753	1,950	525	1,000	2,000	8,000	4,000	0	0	0	0
H20 Revenue	4,600	650	50	10	0	0	0	0	0	0	0	0
RTX Pro 6000 Blackwell Server Edition Revenue	0	0	1,400	1,750	1,050	0	0	0	0	0	0	0
Rubin Revenue	0	0	0	0	0	3,838	26,217	52,481	113,801	117,776	92,184	33,341
Rubin Ultra Revenue	0	0	0	0	0	0	0	0	1,254	38,478	104,981	0
CPX Revenue	0	0	0	0	0	0	0	0	1,483	1,378	2,469	0
Nvidia Groq LP30/35 Revenue	0	0	0	0	0	1,440	2,760	3,600	3,600	3,600	3,600	3,600
DC compute other Revenue	623	639	686	687	678	692	706	720	734	749	764	779
<b>Nvidia Datacenter GPUs revenue reported in Compute</b>	<b>34,155</b>	<b>33,844</b>	<b>43,028</b>	<b>51,334</b>	<b>64,994</b>	<b>77,107</b>	<b>91,887</b>	<b>101,019</b>	<b>132,701</b>	<b>136,336</b>	<b>141,129</b>	<b>149,896</b>
<b>Nvidia NVLink Revenue reported in Networking</b>	<b>1,442</b>	<b>1,523</b>	<b>2,297</b>	<b>2,569</b>	<b>3,481</b>	<b>4,346</b>	<b>4,292</b>	<b>4,589</b>	<b>4,998</b>	<b>5,048</b>	<b>3,654</b>	<b>1,380</b>
<b>Nvidia whole company revenue</b>	<b>44,062</b>	<b>46,743</b>	<b>57,006</b>	<b>68,127</b>	<b>81,741</b>	<b>96,734</b>	<b>113,438</b>	<b>124,069</b>	<b>159,515</b>	<b>164,222</b>	<b>168,442</b>	<b>176,200</b>
% change vs previous												
B100/200/300 Revenue	-	-	-	-	1%	5%	-	-	-	-	-	-
B200A/300A Revenue	-	-	-	-	-	-	-	-	-	-	-	-
B30A Revenue	-	-	-	-	-	-	-	-	-	-	-	-
H100/H200 Revenue	-	-	-	-	(50%)	(67%)	33%	100%	-	-	-	-
H20 Revenue	-	-	-	-	-	-	-	-	-	-	-	-
RTX Pro 6000 Blackwell Server Edition Revenue	-	-	-	-	-	-	-	-	-	-	-	-
Rubin Revenue	-	-	-	-	-	-	-	-	-	-	-	-
Rubin Ultra Revenue	-	-	-	-	-	-	-	-	-	-	-	-
CPX Revenue	-	-	-	-	-	-	-	-	-	-	-	-
LPU Revenue	-	-	-	-	-	-	-	-	-	-	-	-
DC compute other revenue	-	-	-	-	-	-	-	-	-	-	-	-
<b>Nvidia Datacenter GPUs revenue reported in Compute</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>(1%)</b>	<b>(1%)</b>	<b>2%</b>	<b>2%</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>
<b>Nvidia NVLink Revenue reported in Networking</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>1%</b>	<b>5%</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>

Source: SemiAnalysis

SUPPLY	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26	1Q27	2Q27	3Q27	4Q27
<b>18 May 2026 update</b>												
B100/200/300 Units (CoWoS-L)	696,236	771,964	1,004,504	1,253,064	1,540,440	1,725,040	1,348,696	994,504	332,520	262,480	112,504	112,504
B200A/300A Units	0	0	0	0	0	0	0	0	0	0	0	0
B30A units	0	0	130,000	30,000	4,248	0	0	0	0	0	0	0
H100/H200 Units	219,568	227,271	111,400	30,000	50,000	100,000	400,000	200,000	0	0	0	0
H20 Units	454,000	59,528	350,000	0	0	0	0	0	0	0	0	0
RTX Pro 6000 Blackwell Server Edition Units	0	0	200,000	600,000	500,000	0	0	0	0	0	0	0
Rubin Units	0	0	0	0	0	60,608	413,976	830,000	1,603,280	1,657,880	1,297,640	469,304
Rubin Ultra Units	0	0	0	0	0	0	0	0	8,352	256,176	698,944	0
Vera CPU Units	0	0	0	0	0	25,704	175,572	348,228	733,500	758,826	708,525	536,590
Rubin CPX Units	0	0	0	0	0	0	0	0	0	61,357	57,030	102,164
Nvidia Groq LP30/35 units	0	0	0	0	0	0	120,000	230,000	300,000	300,000	300,000	300,000
B100/200/300 Revenue	23,732	27,802	38,942	48,362	62,266	70,577	55,524	41,058	14,566	11,474	4,725	4,725
B200A/300A Revenue	0	0	0	0	0	0	0	0	0	0	0	0
B30A Revenue	0	0	1,950	450	64	0	0	0	0	0	0	0
H100/H200 Revenue	5,200	4,753	1,950	525	1,000	2,000	8,000	4,000	0	0	0	0
H20 Revenue	4,600	650	4,050	0	0	0	0	0	0	0	0	0
RTX Pro 6000 Blackwell Server Edition Revenue	0	0	1,400	4,200	3,500	0	0	0	0	0	0	0
Rubin Revenue	0	0	0	0	0	3,838	26,217	52,481	113,801	117,776	92,184	33,341
Rubin Ultra Revenue	0	0	0	0	0	0	0	0	0	1,254	38,478	104,981
Rubin CPX Revenue	0	0	0	0	0	0	0	0	0	1,483	1,378	2,469
Nvidia Groq LPU Revenue	0	0	0	0	0	1,440	2,760	3,600	3,600	3,600	3,600	3,600
DC compute other Revenue	623	639	686	687	678	692	706	720	734	749	764	779
<b>Nvidia Datacenter GPUs revenue reported in Compute</b>	<b>34,155</b>	<b>33,844</b>	<b>48,978</b>	<b>54,224</b>	<b>67,508</b>	<b>77,107</b>	<b>91,887</b>	<b>101,019</b>	<b>132,701</b>	<b>136,336</b>	<b>141,129</b>	<b>149,896</b>
<b>Nvidia NVLink Revenue reported in Networking</b>	<b>1,442</b>	<b>1,523</b>	<b>2,297</b>	<b>2,569</b>	<b>3,481</b>	<b>4,346</b>	<b>4,292</b>	<b>4,589</b>	<b>4,998</b>	<b>5,048</b>	<b>3,654</b>	<b>1,380</b>
<b>Nvidia whole company revenue</b>	<b>44,062</b>	<b>46,743</b>	<b>57,006</b>	<b>68,127</b>	<b>84,556</b>	<b>96,734</b>	<b>113,438</b>	<b>124,069</b>	<b>159,515</b>	<b>164,222</b>	<b>168,442</b>	<b>176,200</b>
% change vs previous												
B100/200/300 Units (CoWoS-L)	-	-	-	-	1%	5%	-	-	-	-	-	-
B200A/300A Units	-	-	-	-	-	-	-	-	-	-	-	-
B30A units	-	-	-	-	-	-	-	-	-	-	-	-
H100/H200 Units	-	-	-	-	(50%)	(67%)	33%	100%	-	-	-	-
H20 Units	-	-	-	-	-	-	-	-	-	-	-	-
RTX Pro 6000 Blackwell Server Edition Units	-	-	-	-	-	-	-	-	-	-	-	-
Rubin Units	-	-	-	-	-	-	-	-	-	-	-	-
Rubin Ultra Units	-	-	-	-	-	-	-	-	-	-	-	-
Vera CPU Units	-	-	-	-	-	-	-	-	-	-	-	-
CPX Units	-	-	-	-	-	-	-	-	-	-	-	-
Nvidia Groq LP30/35 units	-	-	-	-	-	-	-	-	-	-	-	-
B100/200/300 Revenue	-	-	-	-	1%	5%	-	-	-	-	-	-
B200A/300A Revenue	-	-	-	-	-	-	-	-	-	-	-	-
B30A Revenue	-	-	-	-	-	-	-	-	-	-	-	-
H100/H200 Revenue	-	-	-	-	(50%)	(67%)	33%	100%	-	-	-	-
H20 Revenue	-	-	-	-	-	-	-	-	-	-	-	-
RTX Pro 6000 Blackwell Server Edition Revenue	-	-	-	-	-	-	-	-	-	-	-	-
Rubin Revenue	-	-	-	-	-	-	-	-	-	-	-	-
Rubin Ultra Revenue	-	-	-	-	-	-	-	-	-	-	-	-
CPX Revenue	-	-	-	-	-	-	-	-	-	-	-	-
CPX Units	-	-	-	-	-	-	-	-	-	-	-	-
DC compute other revenue	-	-	-	-	-	-	-	-	-	-	-	-
<b>Nvidia Datacenter GPUs revenue reported in Compute</b>	-	-	-	-	<b>(1%)</b>	<b>(1%)</b>	<b>2%</b>	<b>2%</b>	-	-	-	-
<b>Nvidia NVLink Revenue reported in Networking</b>	-	-	-	-	<b>1%</b>	<b>5%</b>	-	-	-	-	-	-

Source: SemiAnalysis

# ByteDance ASIC VeriSilicon, Alphawave Involvement Possible?

ByteDance has developed an in-house AI accelerator known as SeedChip V1. Chinese IP and design house VeriSilicon is the design services partner and tape-out intermediary, and the chip will be manufactured by Samsung Foundry on SF4. The chip is a monolithic design that is LPDDR based (no HBM).

We see SeedChip V1 ramping Q3 with 120k units total in 2026, and just under 600k units in 2027. We estimate ASP of \$1k to VeriSilicon.

There is also the possibility that AlphaWave is involved and providing SerDes IP for the project, though this is yet to be confirmed. This would also fit with

QCOM's recent earnings call about shipping an ASIC to a hyperscaler. The rationale behind our theory is:

1. Alphawave is the key supplier of high-speeds SerDes IP that is Samsung Foundry qualified. This the same reason why Alphawave is the SerDes IP provider for Groq's LPU v3/LP30.
2. There is (or was) a close relationship between VeriSilicon and Alphawave. VeriSilicon is the official reseller of Alphawave IP in China. There was a formal JV between Alphawave and VeriSilicon for this partnership though this entity ('WiseWave') was put on the entity list before Alphawave exited in order to pave the way for the acquisition by Qualcomm. In addition, Marvell co-founders Weili Dai and Sehat Sutardja were heavily involved in Alphawave as directors and investors. The founder and CEO of VeriSilicon is Wayne Dai, Weili Dai's brother.

Broadcom's work with ByteDance continues to be on hiatus.

## Trainium3 Server Mix

We lower our Trainium3 Teton Max mix in favor of Trainium3 Teton PDS, given Amazon's preference for the PDS version.

## Installed Base

- Add ~4k GB300s to Jane Street
- Add ~120k SeedChip V1 to ByteDance

[Download Model](#)

---

## Disclaimers

## **Analyst Certifications and Independence of Research.**

Each of the analysts whose names appear in this report hereby certify that all the views expressed in this Report accurately reflect our personal views about any and all of the subject securities or issuers and that no part of our compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views in this Report. SemiAnalysis LLC (the “Company”) is an independent equity research provider. The Company is not a member of the FINRA or the SIPC and is not a registered broker dealer or investment adviser. SemiAnalysis has no other regulated or unregulated business activities which conflict with its provision of independent research.

## **Limitation Of Research And Information.**

This Report has been prepared for distribution to only qualified institutional or professional clients of SemiAnalysis LLC. The contents of this Report represent the views, opinions, and analyses of its authors. The information contained herein does not constitute financial, legal, tax or any other advice. All third-party data presented herein were obtained from publicly available sources which are believed to be reliable; however, the Company makes no warranty, express or implied, concerning the accuracy or completeness of such information. In no event shall the Company be responsible or liable for the correctness of, or update to, any such material or for any damage or lost opportunities resulting from use of this data. Nothing contained in this Report or any distribution by the Company should be construed as any offer to sell, or any solicitation of an offer to buy, any security or investment. Any research or other material received should not be construed as individualized investment advice. Investment decisions should be made as part of an overall portfolio strategy and you should consult with a professional financial advisor, legal and tax advisor prior to making any investment decision. SemiAnalysis LLC shall not be liable for any direct or indirect, incidental or consequential loss or damage (including loss of profits, revenue or goodwill) arising from any investment decisions based on information or research obtained from SemiAnalysis LLC.

## **Reproduction and Distribution Strictly Prohibited.**

No user of this Report may reproduce, modify, copy, distribute, sell, resell, transmit, transfer, license, assign or publish the Report itself or any

information contained therein. Notwithstanding the foregoing, clients with access to working models are permitted to alter or modify the information contained therein, provided that it is solely for such client's own use. This Report is not intended to be available or distributed for any purpose that would be deemed unlawful or otherwise prohibited by any local, state, national or international laws or regulations or would otherwise subject the Company to registration or regulation of any kind within such jurisdiction.

### **Copyrights, Trademarks, Intellectual Property.**

SemiAnalysis LLC, and any logos or marks included in this Report are proprietary materials. The use of such terms and logos and marks without the express written consent of SemiAnalysis LLC is strictly prohibited. The copyright in the pages or in the screens of the Report, and in the information and material therein, is proprietary material owned by SemiAnalysis LLC unless otherwise indicated. The unauthorized use of any material on this Report may violate numerous statutes, regulations and laws, including, but not limited to, copyright, trademark, trade secret or patent laws.

### **Private Investment Disclosure.**

Each analyst and/or author contributing to this Report hereby discloses that they may hold, directly or indirectly, private investments in companies, sectors, or asset classes discussed or referenced herein. Such holdings, if any, are disclosed externally with a third-party vendor. SemiAnalysis LLC requires that no author or analyst allow their personal investment positions to influence the views, opinions, analyses, or conclusions expressed in this Report. Readers should be aware that the existence of such holdings, even where disclosed, may present a potential conflict of interest, and are encouraged to weigh this accordingly when evaluating the contents of this Report.