LLNL-TR-815497

Lawrence Livermore National Laboratory

OFFICE OF DEFENSE COORDINATION

China's Bid To Lead Artificial Intelligence Chip Development within the Decade

Mary Gullett and Samantha Bleykhman

9 October 2020

Lawrence Livermore National Laboratory,
P.O. Box 808, Livermore, CA 94551-0808

Disclaimer

Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the US Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.

Executive Summary

In 2017, the Chinese government announced an ambitious, broad strategy to become the global leader in artificial intelligence (AI) theories, technologies, and applications by 2030 and indicated that China's ability to indigenously produce cutting-edge AI chips would be integral to its success. China's goal to become a dominant producer of AI hardware within a decade is a bold undertaking because China faces domestic chip production challenges and fierce competition from US chip producers.

Our review of Chinese AI chip product lines shows various levels where Chinese companies compete with global firms in the production chain. China lacks a robust indigenous AI chip production infrastructure, and the United States and its allies are tightening controls on the supply of advanced semiconductor manufacturing equipment (SME) and electronic design automation (EDA) software to China. In contrast to their Chinese counterparts, US companies operate in all areas of AI chip production and have consistently driven the development of new chip designs.

AI chip production is a costly endeavor, and China's chip producers currently lack the customer base to lead global AI chip development by the 2030 goal. An expanded customer base may help China attract the necessary expertise to bolster its AI chip production chain. This would also help Chinese producers make their investments in the requisite technology economically sustainable. Chinese researchers continue to pursue promising next-generation AI chip designs, such as neuromorphic computing-based architectures, which in the next five to ten years could position Chinese companies to make or contribute to key innovations in the field.

China's market share in AI chip technology could grow in the next few years if China continues to focus on developing low-cost application-specific integrated circuits (ASICs) for AI applications. China's leading chip-making companies heavily focus on ASICs; six of the top ten companies specialized in producing ASICs as of 2018. China has some local fabrication capacity for chips at trailing technology nodes, and ASICs designed at these older nodes can still provide performance advantages over chips designed for a wider range of applications.

Scope Note

This National Security Studies (NSS) report was produced for the Office of Defense Coordination (ODC) at Lawrence Livermore National Laboratory (LLNL) and examines China's ambitions to become a global leader in AI chip technology within a decade. The report evaluates China's role in the global market for AI chips, focusing on market leaders, comparing chip performance across several key metrics, and estimating trendlines going forward—including emerging processor architectures, such as neuromorphic processors, and nascent application areas, such as in space. The report does not include important but tangential information on AI chip hardware-enabled security since companies generally withhold security information as privileged. However, we acknowledge that as chip technology evolves, continual research and development (R&D) into mechanisms for combating infiltration will be necessary for companies to maintain a leading-edge position in the market.

Many definitions exist among different communities for the term "chip." For the purposes of this report, we use the term broadly to refer to a semiconductor-based circuit on a single substrate that usually performs the functions of a processor (in some circumstances, chips may be designed for memory or storage). A (computing) processor is a type of chip; we define a processor as a combination of hardware elements (logical Boolean functions) that perform some combinatory calculations resulting in a computable solution. This report is drawn entirely from company websites and datasheets, news reporting, research centers, technology blogs, and scientific journals. It relies heavily on chip specification sheets available on company websites regarding chip performance.

China Sees Breakthroughs in AI Hardware as Critical to Its Goal to Lead in AI

In a 2017 national strategy document, the Chinese government stated its intent for China to become the global leader in AI theories, technologies, and applications by 2030.[1] In that document, Chinese leaders emphasized that China's ability to indigenously develop intelligent computing chips—what they described as "energy-efficient, reconfigurable brain-like computing chips" and " high-efficiency brain-inspired neural network architectures and hardware systems with autonomous learning capabilities"—would be integral to its success. However, Chinese leaders acknowledged that China lacked original contributions in this area and lagged global leaders in the development of cutting-edge chips for AI applications. In a 2018 report, a leading scholar on China's approach to strategic technologies argued that "[d]ue to their high initial costs and long creation cycle, processor and chip development may be the most difficult component of China's AI plan."[2]

China's ambitions to lead in AI aim to improve China's economic competitiveness and at least in part to overcome what China perceives as a gap between Chinese and US military capabilities.[3] In his October 2017 report to the 19th Party Congress, Chinese President Xi Jinping urged the People's Liberation Army to "[a]ccelerate the development of military intelligentization," according to 2019 testimony before the US-China Economic and Security Review Commission Hearing on Trade, Technology, and Military-Civil Fusion.[4] In 2019, a Chinese military scholar described the concept of "intelligentization" or "intelligentized warfare" as an evolution in warfare in which traditional "human-on-human" military conflicts are superseded by "machine-on-human" or "machine-on-machine" warfare;[5] the concept involves applying AI's machine speed and processing power to military strategy and operations. The term is commonly used by Chinese scholars and has served as a guiding principle for the future of Chinese military modernization. In this context, China's leadership—including President Xi Jinping—views AI as a technology that could provide China with a decisive strategic advantage.

China Faces Economic and Technological Hurdles in Its Path to Self-Sufficient AI Chip Production

China's longstanding dependence on foreign suppliers for advanced processors, despite China investing billions of dollars over several decades to create a domestic semiconductor industry (**figure1**),[*,6] suggests that it is likely to face challenges in its bid to become a leading global producer of AI chips. The steep up-front investment costs and performance requirements of AI chips have hindered China's ability to break into the market in a significant way. (For an overview of why AI chips must be advanced and specialized, see the **appendix** on page 19.) Nevertheless, China continues to invest heavily in these efforts. In 2014, Beijing established the National Integrated Circuits Industry Investment Fund, a subsidy program that plans to raise US$180 billion from local-government-backed funds and state-owned enterprises.[7] In 2015, China announced its "Made in China 2025" initiative, a program aimed at upgrading China's entire manufacturing industry that set a goal for China to produce US$305 billion worth of chips annually and meet 80 percent of domestic demand for chips by 2030.[8]

[*] In broad terms, a semiconductor is a material that conducts electricity more than an insulator but less than a pure conductor. Modern chips are made from semiconductor material. A 2019 report issued by the Center for Strategic and International Studies (CSIS) (reference 6) notes that China's State Council in 2014 set a goal of becoming a global leader in all segments of the semiconductor industry by 2030.

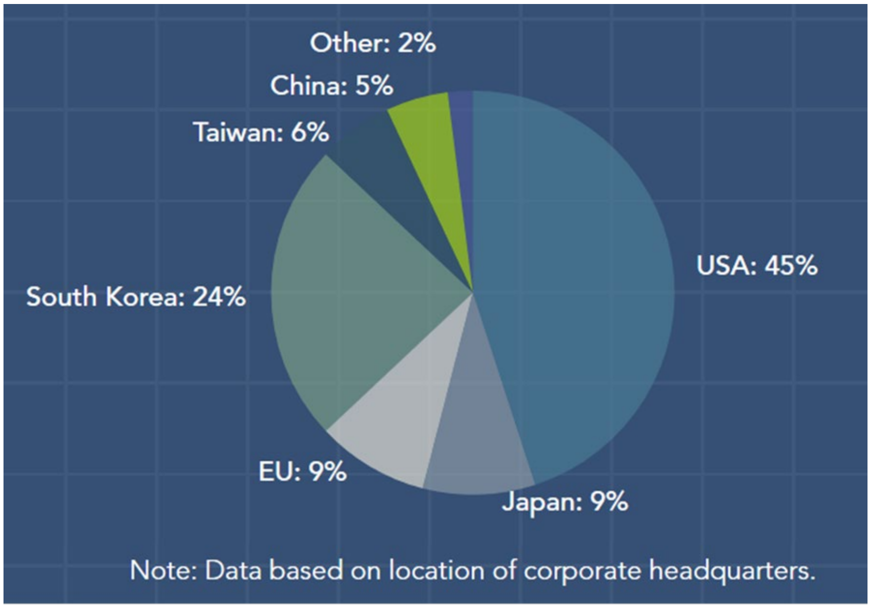

Figure 1. Market share in the global semiconductor industry as of 2018, as shown by the Semiconductor Industry Association.[9] (EU = European Union)

AI Chip Production Is a Costly Endeavor…

These continued investments will be essential if China is to be successful. Reliable estimates suggest that it costs around US$10 to $15 billion to build a chip manufacturing plant capable of producing the most advanced processors—meaning processors with a node size of less than 7 nanometers (nm).[10,11] These high costs have sometimes discouraged even Western companies from pursuing leading nodes; in 2018, a leading US foundry abandoned plans to build a 7-nm fabrication plant, citing high costs.[12]

A key factor in these high costs is the equipment and machinery needed to fabricate the chips. The technique used to imprint circuitry patterns onto wafers is photomasking, a process that requires photolithography machines—one of the most expensive tools in the fabrication process.[13] A single extreme ultraviolet photolithography machine from the Dutch company ASML (the main international supplier of these machines) in 2019 cost US$172 million, according to a Bloomberg report.[14] A plant with a sizeable production line would require multiple, perhaps even dozens, of these machines.[15]

…and Requires Deep Technical Expertise

While continued investment will be critical for China, spending alone will be insufficient because of the deep technical expertise required at multiple stages of AI chip production. The key steps necessary to produce an AI chip are (1) processor production, (2) chip design, (3) fabrication, and (4) assembly and packaging. Steps 1 and 4 resemble the processes used for producing other types of semiconductors:

- When a chip vendor creates designs for its chips (step 1, processor production), it typically incorporates its own custom AI processor with graphics processing units (GPUs) and/or central processing units (CPUs)[*] from other companies; these base processors for the most part are sufficiently available that they do not constitute a major bottleneck in the production chain.

[*] GPUs and CPUs are microprocessors that use different architectures to serve different purposes. A CPU is suited to a wide variety of tasks, such as serial computing and running databases. GPUs, by contrast, use highly parallel architectures to carry out simultaneous operations. For more detail, see the appendix.

- Step 4, assembly and packaging, involves encasing chips in a supporting case (typically ceramic or plastic) to protect them against physical damage and corrosion. As with step 1, this step is common among all types of chips, not just those designed for AI applications.

The key chokepoints in the production chain are in steps 2 and 3, chip design and fabrication:

- The R&D necessary to make advances in chip design can span decades.[16] US companies dominate the market for EDA software,[*,17] and Chinese companies rely heavily on the EDA software from these companies to design their chips. For example, Cambricon—a state-owned Chinese company that produces chips for intelligent cloud servers, terminals, and robots—licenses its designs from the Silicon Valley chip designer Arteris for the backbone of the interconnects that move data around its chips, according to a 2017 *Wired* article.[18] In 2019, Technode—the Chinese partner of TechCrunch, a US online publisher focused on the technology industry—provided four reasons for China's EDA gap with the United States: (1) Chinese EDA tools are not comprehensive; (2) Chinese EDA firms lack the engineering expertise to develop advanced EDA software (the article notes that only about 300 of China's 1,500 EDA specialists work for Chinese firms); (3) market entry is difficult; and (4) Chinese companies do not have enough access—meaning access to the leading chip foundries—to keep up with relevant developments in manufacturing.[19]
- Meanwhile, fabrication of a leading-edge AI chip requires expertise at the forefront of physics, materials science, and micromechanical engineering—e.g., making mechanical devices out of a process that essentially involves one atom at a time and incorporating an increasing amount of power, memory, and functions onto a single wafer.[20] Although China has proven adept at reverse-engineering, reverse-engineering a state-of-the-art AI chip fast enough for it to be leading-edge by 2030 is difficult. As a senior official at the Semiconductor Industry Association noted in a 2019 Quartz article, "…[L]et's say you've got 5,000 engineers that can reverse engineer the chip and look at the chip layout—by the time you've done that, the American company's already two generations ahead of you."[21]

Figure 2 is a flow chart of the AI chip production chain that identifies the two major chokepoints. The next section addresses China's difficulties with production steps in more detail.

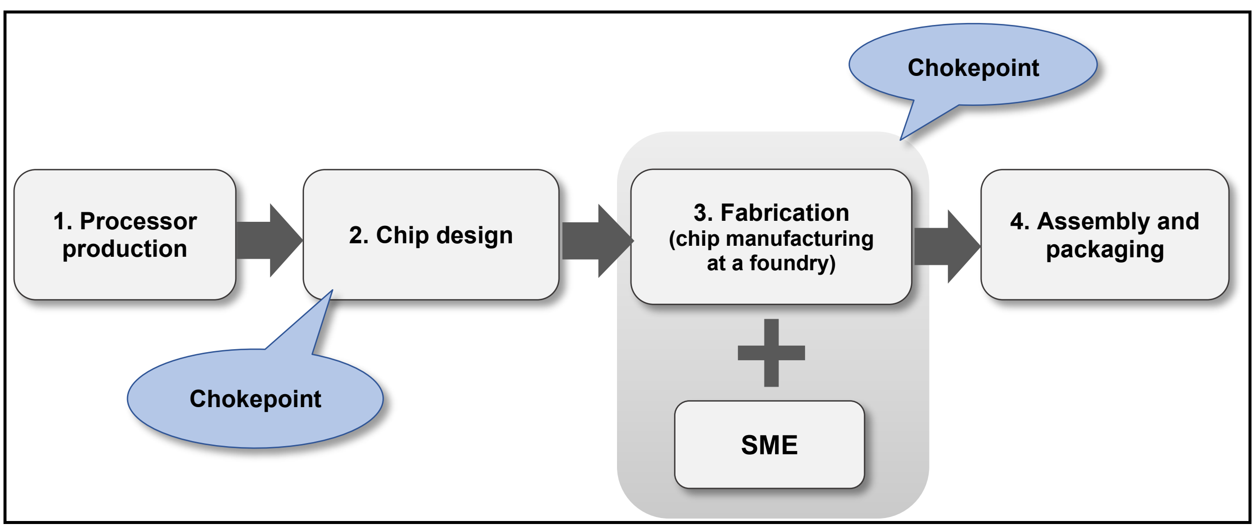

Figure 2. Key steps and chokepoints in the AI chip production chain.

[*] EDA tools are used to design integrated circuits; since modern chips can incorporate billions of components, these automation tools are essential for their design.

China Struggles To Compete with United States and Other Western Companies in Key Steps of the Production Chain

Chinese companies are not global leaders in any of the production steps and have struggled to break into most production steps—and the production of SME, in particular—in a significant way. By contrast, US companies as of 2020 operate in all stages of the AI chip production chain, and several are market leaders in processor production, chip design, SME production, and chip sales. (**Figure 3** identifies global leaders in the AI chip production chain based on the country in which the company has its headquarters.) Although China's Pingtouge (an Alibaba subsidiary) and HiSilicon (a Huawei subsidiary) are engaged in processor production, both companies outsource their chip design (whereas many companies design their chips in-house). Some Chinese companies have design capabilities—for example, Huawei's Kirin 980 AI system-on-a-chip (SoC) incorporates Chinese-designed neural network processing units (NPUs), according to the chip's specification sheet available on the company's website[22] and a product review.[23] However, these chips to date have mostly been designed for niche market areas, such as cellular devices. The US companies Nvidia and AMD dominate the international design market for GPUs, while China's leading GPU company, Jingjia Microelectronics, fields GPUs that are significantly slower.[24]

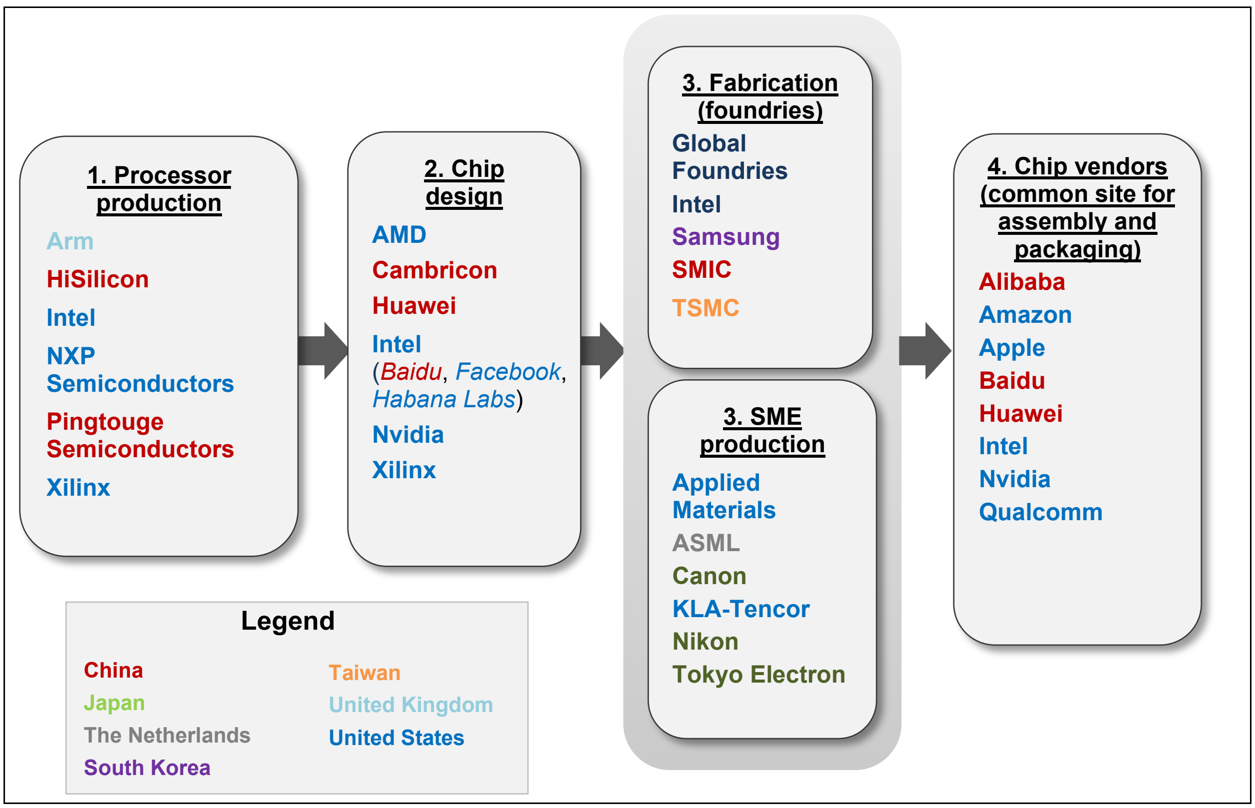

Figure 3. Global leaders in the AI chip production chain. Numbers indicate steps 1 to 4 as identified in the previous section. Companies are listed in alphabetical order. Country association is based on where the company is headquartered.

Producing SME is one of China's most significant technological hurdles. The United States, Japan, and the Netherlands lead the market in SME, while China's capabilities remain nascent. (**Table 1** shows a table of SME providers for AI chip foundries.) The Dutch company ASML is known for producing the most leading-edge machinery, capable of fabricating processors at node sizes of less than 7 nm (with a smaller node size, more transistors can fit onto the chip, thus increasing the transistor density). Only two Chinese companies operate in the SME market segment—the Advanced Microfabrication Equipment Company (AMEC)[25] and Wuhan Hongxin Semiconductor Manufacturing (HSMC)[26]—and they have fewer product lines than their international competitors. AMEC claims that its etching tools are capable of fabricating processors with a node size of 5 nm,[27] but its main production line is at 7 nm and above.[*,28,29] HSMC, for its part, was established in late 2017 and produces equipment for manufacturing processors at the 14-nm level; it is conducting R&D of 7-nm technology, aiming to begin testing chips at that size in late 2021.[30]

Table 1. SME providers for AI chip foundries. AI chip fabrication requires many types of complex machines and processes. We focus on wafer etching and testing machines because they directly impact node size fabrication capabilities. Other SME machinery includes fabrication machinery, wafer preparation and packaging machinery, and deposition equipment.

Company	AMEC[31]	HSMC[32,33]	Ushio[34]	Canon[35,36]	Nikon[37]	Tokyo Seimitsu[38]	Tokyo Electron[39]	ASML[40,41]	KLA-Tencor[42]	Lam Research[43,44]	Applied Materials[45,46]
Country	China	China	Japan	Japan	Japan	Japan	Japan	Netherlands	US	US	US
Machine name/ series	Primo D-RIE series	14/7nm lines	EUV light source	KrF -- FPA 6300	ArF/KrF -- NSR line	WGM 5200	Certas LEAGA	NXT & NXE series	ATL 7nm UV line	UV—SOLA line	EUV Centura-Tetra, 14nm
Machine type	Fabrication	Fabrication	Light source	Fabrication	Fabrication	Fabrication and testing	Fabrication, etch, deposition	Fabrication	Fabrication, etch, deposition	Fabrication, testing, etch, deposition	Fabrication and testing, wafer prep, etch machines

Taiwan's TSMC dominates the international foundry market for AI chips, controlling about 70 percent of the market share for training, inference, cellular, and cloud/edge chip manufacturers, based on our review of AI-chip vendor product data sheets. TSMC also controls more than half the market for all types of chip production, according to a 2019 Bloomberg report.[47] Other foundries occupying a share of the market are based in the United States and South Korea. China and Japan also operate in-country foundries, but none of the chips we researched were identified as being manufactured in those facilities. **Figure 4** identifies node-size fabrication capabilities as of 2020 for companies based in the United States, China, Japan, South Korea, and Taiwan, and it shows South Korea's Samsung and Taiwan's TSMC leading US companies by a generation and China's SMIC by two generations. However, a 2020 overview of AI chips by the Center for Security and Emerging Technology (CSET) notes that SMIC's production capacity at the 14-nm level is limited and that SMIC relies on SME imports from the United States, the Netherlands, and Japan.[48] In September 2020, the Department of Commerce announced that it would require US chip companies to obtain licenses before exporting technology to SMIC because exports to the company or its subsidiaries risk being used for Chinese military activities, according to US press reports.[49,50] US market analysts predict that the decision will degrade the quality and yield of SMIC's chip production.[51] Both Samsung and TSMC have plans for 3-nm fabrication around 2022, making it difficult for competitors to keep pace with these two companies.

---

[*] In 2018, the Dutch government gave ASML a license to sell its most advanced machine to a Chinese customer, according to a Reuters report (reference 29). ASML never publicly disclosed the identity of the Chinese customer, but other media outlets reported that it was Semiconductor Manufacturing International Corp (SMIC), China's largest chip-making specialist. The Dutch government later decided not to renew the license after facing pushback from US government officials.

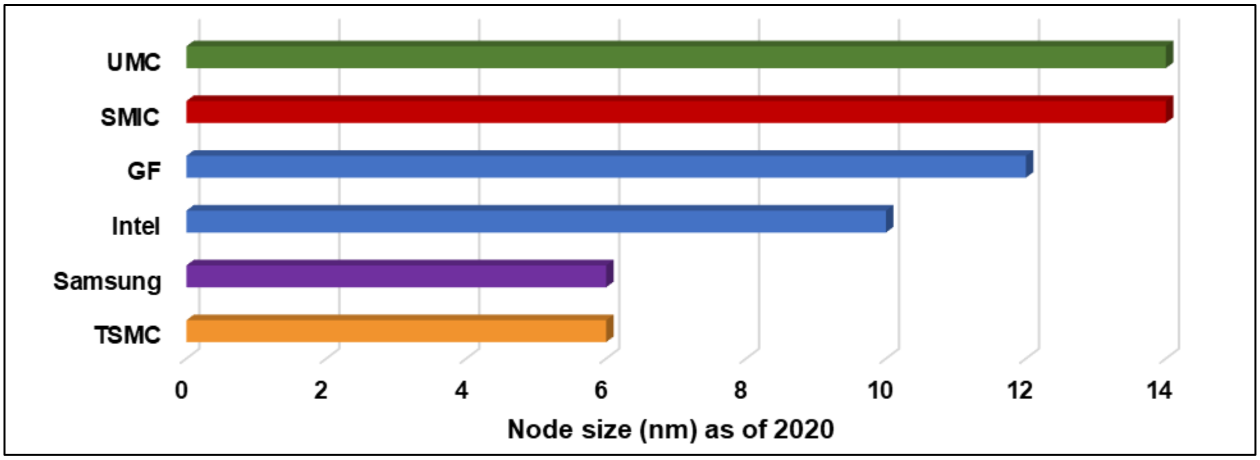

Figure 4. Bar plot identifying the most advanced node size that each foundry uses. Taiwan's TSMC and South Korea's Samsung lead the foundry industry when it comes to node-size capability; they lead China's SMIC by 2 generations. As Samsung and TSMC have plans already set for 3-nm fabrication around 2022, competitors struggle to keep pace with these two companies.[52,53,54,55,56,57,58,59]

Chinese AI Chips Are Generally Designed for Niche Market Areas

AI chips offered by Chinese companies are not widely competitive outside of niche application areas. Although US companies do not lead the market in each step of the AI chip production chain, they possess capabilities in each production step. They also lead the market in global AI chip sales, offering the most versatile and high-throughput AI chips in the market. Chips offered by US companies are applied across essentially all industries that leverage machine learning (e.g., automotive, healthcare, retail, telecommunications). By contrast, AI chips offered by Chinese companies are not as widely proven across the industry. We compare leading global and Chinese-origin AI chips in the sections that follow according to their application type.

Training and Inference (T/I) AI Chips: Gaps in Information Reporting Hamper Direct Comparison

A comparison of leading US and Chinese chips in this variety based on the chips' specification sheets is challenging because there are multiple gaps in the reported performance parameters for Chinese T/I chips and there may be anomalies in the reported data (**table 2**). For example, a lack of reported information on the throughput and efficiency of Huawei's Ascend 910 chip and Enflame Tech's CloudBlazer T11 makes it challenging to compare Chinese and US chips across these metrics. The precision performance indicators (Mixed Precision FP32 and INT8 Performance) for the Chinese chips shown in Table 2 could be an anomaly because the data are so far outside the range of the reported values for comparable US chips. Blank cells are those for which the information was not available; purple boxes highlight groupings of such gaps.

In the training chip variety, Nvidia has been the longstanding leader, and a 2019 Bloomberg article projected that Nvidia would continue to dominate "for some time because of their robust and flexible software platform and stickiness with AI developers."[60] As seen in table 2, Nvidia appears to provide more in-depth information about the performance parameters of its T/I chips than most of its competitors.

Table 2. Listed performance of Chinese and US T/I chips. Purple boxes highlight gaps in reported parameters for the Chinese chips. (DTU = deep thinking unit; MHz = megahertz; W = watt; Gbps = gigabit per second; Tbps = terabit per second; GB = gigabyte; TB = terabyte; s = second; TOPS = tera-operations per second; TFLOPS = tera-floating-point operations per second; IPS = inferences per second.)

T/I chip comparison						
	Huawei Ascend 910[61,62,63,64]	Enflame Tech CloudBlazer T11[65,66,67]	Nvidia T4 (RTX 2080)[68,69,70]	Nvidia V100[71,72]	Intel GOYA[73,74]	Intel GAUDI[75,76,77,78]
	China	China	US	US	US	US
Accelerator chip type	CPU	DTU	GPU	GPU	CPU	CPU
Chip release date (month/year)	8/19		9/18	9/17	1/18	6/19
Fabrication process (nm)	7	12	12	12	16	16
Fabrication provider	TSMC	Global Foundries	TSMC	TSMC	TSMC	TSMC
Transistors			13.6 billion	21.1 billion		
Processor model			TU104	GV100	HL-1000	HL-2000
Processor/core clock (MHz)			585/1,515	1,530		
Power draw, TDP (W)	310	300	70/215	300	100	200
Memory base clock (MHz)			1,250	876		
Data transmission rate	240		10/14 Gbps			2 Tbps
Memory size (GB)	16/32	16	16	16/32*		32
Peak memory bandwidth	1.2 TB/s	512 (GB/s)	320/448 (GB/s)	900* (GB/s)		1 TB/s
Mixed precision FP32 (TFLOPS)	**128**	22	8.1/10	15.7		
INT8 performance (TOPS)	**512**	n/a	130	n/a	n/a	
Throughput (ResNet-50, IPS)			4,018	6,275	15,000	3,300
Peak efficiency (ResNet-50, IPS/W)			60	22/27	150	16.5

* The V100 attains 32 GB memory with much higher bandwidth when GPUs are unified to share memory capacity.

Cellular AI Chips: China's Huawei Competes with US Companies

Our review of AI chips optimized for cellular devices shows that Huawei is the Chinese market leader, while the US companies Apple, Google, and Qualcomm are all competitive in this space. All the companies we researched in this category depend on Taiwan's TSMC 7-nm process for their chip fabrication. The smaller node size is particularly important to this market segment because a small chip size is critical for handheld devices and enables companies to provide the most efficient cellular phone performance. These chips are optimized for computer vision, as cellular phone companies tend to differentiate their products based on phone screens. A phone that processes on the device (rather than routing through the cloud) reduces latency, thus improving the user experience. **Table 3** compares US and Chinese cellular device AI chips across key performance parameters with performance parameters of interest highlighted with a purple box.

Table 3. Listed performance of Chinese and US AI chips for cellular devices. All the newer, high-tech cellular device chips that we evaluated claimed to use a 7-nm fabrication process and Taiwan's TSMC for fabrication. We call attention to the much larger memory in Qualcomm's Snapdragon chip (16 GB of on-device random access memory [RAM]) compared to other leading cellular AI chips (which have storage within the 4- to 8-GB range). We also call attention to Google's use of Qualcomm GPUs in its cellular chip design.

Cellular device chips						
	Huawei Kirin 980 Octa-core[79,80]	Huawei Kirin 990 5G[81,82]	Qualcomm Snapdragon 865 5G[83,84,85]	Apple A12 Bionic[86,87,88,89]	Apple A13 Bionic[90,91,92,93,94]	Google Visual Core[95]
	China	China	US	US	US	US
Phone implementation	Mate-20 Pro	Mate-30 Pro	Galaxy S20	iPhone XR	iPhone 11	Pixel 3/4
Release date	Aug 18	Sept 19	exp. 20	Sept 18	exp. Sept 20	Oct 19
Fabrication process (nm)	N7P	7+, EUV	N7P	N7	N7P	
Fabrication provider	TSMC	TSMC	TSMC	TSMC	TSMC	
Transistor count (billions)	6.9	10.3		6.9	8.5	
Network connectivity		5G	5G	LTE	LTE	LTE
On-device RAM (GB)	6	8	16	4	4	6
CPU	Arm Cortex A76/A55 processors	2 Arm Cortex processors: A76, A55	Qualcomm Kryo 585, Arm Cortex A77/A55	6 Apple CPU cores: 2 Vortex, 4 Tempest	6 Apple CPU cores: 2 Lightning, 4 Thunder	Qualcomm Kryo 485 with Arm Cortex processors
Peak download speed (Gbps)	1.7	2.3	7.5*/2.5	1	1.6	1.2
Upload speed (Gbps)	1.4	1.25	3*/0.31	0.15		0.15
ML performance (TOPS)			15*/10	5	5	
Sub-6-GHz bandwidth (MHz)	160		100			
GPU	Arm Mali-G76 processor	Arm Mali-G76 processor	Qualcomm Adreno 650	Apple A12 Bionic Neural Engine	Apple A13 Bionic Neural Engine	Qualcomm Adreno 640 Octa-core
Processor clock (GHz)	2.6	2.86	2.84	2.49	2.65	

* Indicates a possible anomaly between values publicized by the companies (first value) and reported results from blogs (second value).

Cloud/Edge Computing AI Chips: Performance Specifications Are Often Sparse or Vague

We include cloud platforms as a separate market segment because the interconnected chips used in the networks are optimized to process vast amounts of data—a useful configuration for machine learning applications. Cloud industry leaders Amazon, Google, and Microsoft provide options for different types of computing plans depending on how much customers want to use their accelerator networks (the more processors the customers use, the more they pay). This model has proven successful and helped cement these companies' leading positions in the industry. Alibaba is probably the leading Chinese company in this market segment and claims to have developed cloud platforms with capabilities comparable to US competitors, but it has not made its cloud technology available for commercial use, which makes it difficult to verify the company's assertions. The Chinese company Cambricon released a cloud/edge computing chip in 2019, but it does not appear to be as commercially popular as its competitors. Because there is essentially no commercial market for individual cloud computing AI chips, many performance specifications for the networks of accelerator chips that power cloud systems are unavailable for open or public viewing. Where such information is available, the specification sheets tend to be vague and short.

Compared to other AI chip application categories, the cloud/edge AI chip market segment incorporates a broader variety of chip types that range from ASICs to custom processors (e.g., the -PU variants; see **table 4**). This is because cloud computing often requires companies to customize their chip networks due to greater demands for higher bandwidth and storage capacity, as well as TOPS performance. Cloud/edge AI chips also tend to be more powerful than chips used for T/I and cellular applications. The emergence of South Korea's Samsung as a competitor with Taiwan's TSMC in fabrication agreements is notable in this segment, especially among the Chinese companies.

Table 4. Listed performance of Chinese and US AI chips for cloud/edge computing. However, the data sheets for cloud/edge AI chips are short and vague compared to those used for other applications. (TPU = tensor processing unit; FPGA = field programmable gate array; IPU = intelligence processing unit; TDP = thermal design power.)

Cloud/edge computing chips and platforms						
	Google Cloud TPUv3[96,97,98]	Amazon Inferentia[99,100]	Microsoft Azure Virtual Machines[101,102]	Alibaba Hanguang 800[103,104]	Baidu KUNLUN[105]	Cambricon MLU270[106,107]
	US	US	US	China	China	China
Accelerator chip type	TPU	ASIC	FPGA	CPU/NPU	XPU	IPU
Chip release date (month/year)	1/19	12/19	8/17	9/19	Due 2020	5/19
Fabrication process (nm)				12	14	16
Fabrication provider				Samsung	Samsung	TSMC
Power draw, TDP (W)				160	150	
Data transmission rate (Gbps)		25/100*	25/200			
Memory size (GB)	128	min 8, max 192*	480			
Peak memory bandwidth (GB/s)				400	512	
Mixed precision FP16 (TFLOPS)	840	64				
INT8 performance (TOPS)	n/a	128			260	128
Peak efficiency				500 IPS/W		

* Amazon provides different Inference machine options by linking Inferentia chips together. The smallest configuration is 1 chip (min), and the largest (max) is 16 chips.

Chinese Researchers Are Focusing on the Development of Next-Generation AI Chip Technologies To Demonstrate China's Leadership in AI

Chinese entities are pursuing cutting-edge concepts simultaneously with their efforts to improve and expand the country's overall semiconductor production, seemingly trying to strike a balance between developing infrastructure in areas where they are behind and pursuing some promising "leap-ahead" technologies. They are making noteworthy achievements in cutting-edge technologies; in 2019, the respected scientific journal *Nature* featured the "Tianjic chip" as its cover story. The Tianjic chip is a China-led combined research effort by Tsinghua University, Beijing Normal University, Singapore University of Technology and Design, and the University of California at Santa Barbara to develop a computing platform that would combine nonspiking artificial neural network and neuromorphic architectures (**sidebar**).[108]

- According to *Nature*, the research team demonstrated that the 28-nm chip they developed could achieve 1.6 to 100 times better throughput and 12 to 10,000 times better power efficiency than an Nvidia Titan-Xp GPU.[109] The research team told Chinese media that they expect to complete the R&D stage in 2020 and that the chip will eventually be deployed in autonomous vehicles and smart robots.[110]
- The title of the article characterized the innovations inherent in the chip's design as presenting a potential pathway toward artificial general intelligence—a more expansive form of AI capable of performing virtually any task a human might undertake. While this projection may have been premature, China almost certainly sees the country's ability to lead the development of the next generation of AI chips as a key pillar in its plan to become the global leader in AI by 2030.

The Tianjic chip remains in the R&D phase, whereas both Intel and IBM have developed operational processors. IBM is currently at the forefront of this technology; it entered the neuromorphic chip market earlier than its competitors and has made steady improvements to its TrueNorth processor, which is now capable of simulating 1 million neurons. Internationally, China's Tsinghua University and Russia's MotivNT are pursuing neuromorphic processors, but these countries' efforts remain in the R&D phase (**table 5**).

Table 5. Listed performance of Chinese and US neuromorphic AI chips. US companies IBM and Intel lead neuromorphic chip development; China and Russian companies remain in the R&D phase. (fps = frames per second; GSOPS = giga-synaptic operations per second.)

Neuromorphic AI chips				
	Tsinghua University Tianjic[111]	MotivNT Altai[112]	Intel Loihi[113,114]	IBM TrueNorth[115,116,117,118,119]
	China	Russia	US	US
Chip release date	R&D	R&D	2018	2014
Fabrication process (nm)	28	28	14	28
Fabrication provider		TSMC	Intel	Samsung
Transistors (billion)			2.07	5.4
Neurons	40,000	131,000	128,000	1 million
Synapses (million)	10	67	128	256
Neural cores	156	256	128	4,096
On-chip memory, SRAM (million bits)				428
Power draw (milliwatts)	950	500		70
Throughput (fps)		1,000/2,200		1,200/2,600
Efficiency (fps/W)				6,100
Synapse performance (Synapse ops, GSOPS)	650	67		58
Synapse efficiency (GSOPS/W)	684	134		829
Neuron scalability		67 million	8.3 million	1 billion
Synapse scalability				256 billion

Emerging AI Chip Architectures and Application Areas

While GPUs, field-programmable gate arrays (FPGAs), ASICs, and SoCs dominate the AI chip market today (see the appendix for a more detailed overview of these different chip types), researchers are pursuing alternative methods and architectures to produce systems that can continue to deliver improvements in speed and efficiency. Quantum computing is one such approach—e.g., the notion that because subatomic particles can exist in more than one state at any given time, you could use a quantum computer to exponentially accelerate computations and store substantially more information using less energy.[120] Another approach that appears promising is neuromorphic computing. The chips—or computational building blocks—within neuromorphic computing systems are logically analogous to neurons in the human brain.[121] Several companies are pursuing R&D of neuromorphic chips as an alternative to conventional architectures and power-hungry GPUs.[122]

Neuromorphic Chips Hold Promise for Revolutionizing AI's Real-Time Capabilities

A neuromorphic chip models the structure of a human brain to transmit signals throughout its processor network. Rather than depending on a processor and memory clocks for signal transmission, once a sensor processor receives a signal that obtains a certain threshold value, the input causes a cascade of signal relays among interconnected processors. While the design of these chips is in its early stages, this architecture has the potential to dramatically improve computational efficiency and enable the chip to operate at nearly real-time sensing.[123]

The design and structure of neuromorphic chips are fundamentally different from the AI chips discussed earlier in this report. While both categories of chips are designed to process artificial neural networks and offer performance improvements compared to traditional CPUs, neuromorphic chips are designed for special neural networks called "spiking neural networks (SNNs)."[124] As Intel explains in an overview of the company's neuromorphic computing research on its website, "[e]ach 'neuron' in the SNN can fire independently of the others, and in doing so, it sends pulsed signals to other neurons in the network that can directly change the electrical states of those neurons. By encoding information within the signals themselves and their timing, SNNs simulate natural learning processes by dynamically remapping the synapses between artificial neurons in response to stimuli."[125] This architecture could provide a means to achieve nearly real-time capability because the design would enable the AI to move toward the fastest reaction and calculation time of a system.

Growing Role for AI Chips in Enabling Spacecraft Operations

The chip innovations that are improving the performance of terrestrial systems are also increasing the computational power available to spacecraft.[126] AI technologies can meaningfully improve spacecraft operations, from enabling precise automated control to facilitating on-board tasks, such as image processing or navigation.[127] However, implementing AI hardware on an embedded system introduces size, weight, and power (SWaP) constraints. The harsh and remote environment of space imposes additional requirements on top of these challenges, such as the need for chips to be radiation-resistant, robust, and extensively tested and verified.[128] (Radiation damages the hardware either through its cumulative effects [total ionizing dose, or TID] or through single event effects. Radiation hardening can enable a chip to withstand such errors, but hardened components are generally twice as slow and many times more expensive than components designed for terrestrial applications.)

AI cloud networks are currently used to alleviate information processing on the ground. For example, the U.S. Aerospace Corporation has developed an on-platform AI cloud network called "Space Cloud" that can operate on satellites in orbit, according to the company's website.[129] Space Cloud uses AI and cloud technology to enable satellites to detect and transmit only meaningful data back to Earth—as explained on the website, the AI "can tell the difference between goats and boats."[130] Space Cloud was built with commercially available technology—specifically, an Intel Movidius processor and an extension of Google's open-source Kubernetes cloud tool.[131] The Movidius processor is a specialized processor designed to deliver high-performance machine vision at low power levels and for small form factors (detailed specifications on the Movidius processor are available on Intel's website[132]). Using the Kubernetes cloud, the AI algorithm can adapt to detect different objects based on their location—e.g., search for goats over land and boats over water—using the same satellite.[133]

Satellite companies are pursuing AI chips that can be physically deployed onboard spacecraft in the next few years, but this capability is currently limited. Although several companies produce radiation-hardened processors, these chips do not appear to be optimized for AI applications. As an example, the US company Xilinx produces a space-grade FPGA that is optimized for sensor processing in space—the Virtex-5QV—but the FPGA is not specially designed for deep-learning–based training or inference functions.[134]

For short-duration missions in low Earth orbit (LEO) where the radiation effects are less intense than they are at higher orbits,[*] commercial off-the-shelf (e.g., unhardened) AI chips may perform adequately to meet mission requirements. The European Space Agency (ESA) in 2018 subjected Intel's Movidius Myriad 2 AI processor to radiation at the European Organization for Nuclear Research in preparation for a mission in 2020 in which the ESA is seeking to demonstrate how onboard AI can improve the efficiency of sending Earth observation data back to Earth.[135,136] (The Movidius Myriad 2 AI processor is a low-power SoC that is optimized for drones and virtual reality devices but is not hardened by design to withstand the radiation environment in space [**figure S1**].[137,138]) The mission has been characterized as the first Earth observation satellite with embedded AI.[†,139,140] However, while the Movidius processor may be able to perform as required for a short-duration demonstration mission, its current design is unlikely to meet the requirements of a longer-term strategic mission.

Maximizing AI performance for many future space-based remote sensing applications—especially if the trend toward smaller spacecraft continues—will require memory and throughput optimization, along with near-real-time processing. With further evolution of the technology, neuromorphic chips appear to be most well-suited to meeting these performance requirements.

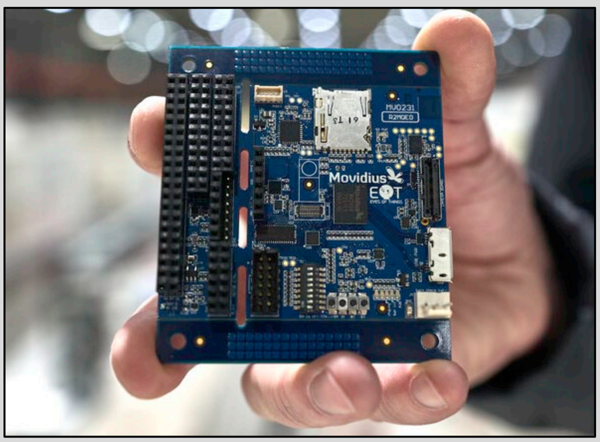

Figure S1. Intel's Myriad 2 processor, requiring less than a watt of power, is designed for SWaP-constrained platforms.[141]

[*] In general, processors should be tolerant to 10 to 50 kilorads (krad) TID for LEO missions and 100 to 300 krad for missions in medium earth orbit or highly elliptical orbit.

[†] The AI technology was deployed on ϕ-Sat, or PhiSat—one of two CubeSats that comprise the FSSCat mission. A hyperspectral camera on one of the CubeSats is collecting an enormous number of images of Earth, some of which will not be suitable for use because of cloud cover. To avoid downlinking the occluded images back to Earth, the ϕ-Sat chip will filter them out so only usable data is returned. For more information about the FSSCat mission, see, for example, references 139 and 140.

China Needs to Address Key Shortcomings in its AI Chip Production Capabilities

China's ability to lead AI chip development by 2030 could depend on whether or not Chinese chip developers invest in the technology required to fill gaps in China's AI chip production chain. This will become particularly important if the United States and its allies further restrict the supply of SME and EDA software to China in the coming years. Because of the high costs and technological barriers, it will require Chinese companies to develop a significant enough domestic and foreign customer base to attract the necessary expertise and make the investment economically sustainable. Developing a foreign market may be complicated by the fact that many Western companies are leery of purchasing Chinese chips out of concern that they may contain backdoors or other security vulnerabilities[142,143]—this distrust is another factor China will need to overcome.

To establish a niche in the AI chip market by 2030, China may emphasize specialized ASIC development. Many Chinese AI chip companies focus on chips optimized for specific applications. For example, Chinese startups Horizon Robotics and Deephi, and the much larger Huawei, have focused on developing chips to bring AI functions such as understanding video to devices such as cars and cameras. While AI chips require advanced designs, if fully optimized for a specific application, many applications can be achieved using lower-cost, older technology nodes.

Appendix: Cutting-Edge AI Applications Require Advanced Chips with Specialized Designs

The chips used to exploit advances in AI are designed to handle data in a fundamentally different way from the CPUs that have been used to execute computing functions for the past several decades. Cutting-edge AI applications require chips that are more powerful, more efficient, and optimized for machine learning computations. A recent publication by CSET provides a comprehensive overview of the evolution from increasingly performant CPUs—meaning CPUs that incorporated steady improvements in transistor-switching speeds and increased transistor density to improve speed and efficiency—toward new chip designs as Moore's Law[*,144] slowed.[145] As long as CPUs were making rapid, steady improvements in speed and efficiency, it was not cost-effective for companies to invest in the design and production of specialized chips.[146] Once improvements in CPUs began to slow around the early 2010s, however, and specialized chips could have longer useful lifetimes, it became more economical for companies to produce them.[147] This trend toward increased computational horsepower and more specialized chip designs coincided with the proliferation of both inexpensive digital sensors (such as smartphones and digital cameras) and data sharing platforms (such as Facebook, Flickr, and YouTube), which have produced massive amounts of training data. These trends in combination have enabled neural network architectures—which are the backbone of deep learning—to exceed human expert performance in several challenging applications (such as object detection in imagery,[148] speech recognition,[149] and moderately difficult video and board games[150,151]), fueling the recent fervor around AI.

AI Chips Are Designed To Perform Large-Scale Calculations for Rapid Data Analysis

AI chip designs vary depending on the specific type of AI application for which the chip is designed. Some common AI applications include training (e.g., training an AI algorithm on a dataset); inference (e.g., performing an inference based on the learned information); cellular devices; and cloud and edge computing. Although these diverse applications require chips with different specific designs, they all share a need to perform large-scale calculations for rapid data analysis based on learned rules that are characteristic of machine learning.

AI computation shares many characteristics with traditional computation but has novel characteristics, as well. As CSET notes, the most significant adaptation of AI chips over traditional CPUs is a parallel computation paradigm.[152] The computations needed to train an AI model are parallelizable because they are identical and not dependent on the results of other computations. GPUs are the preferred chips for this task because their main attribute is parallel processing. By contrast, a traditional CPU is designed to perform serial computations (meaning computations performed sequentially because the result of one computation is required for the next). For other AI functions, such as inference, the applications are more diverse, requiring chips that are customized for particular functions.[153] Further, the parameters for AI computing are generally very large, requiring tremendous storage capacity, high bandwidth, low latency memory access capacity, and robust yet flexible connections between computing units and memory devices.[154]

[*] Moore's Law refers to an observation made by Gordon Moore, a former CEO of Intel, in a 1965 paper (reference 144) that the number of components per integrated circuit doubled every year and his projection that this rate of growth would continue for at least another decade. He later revised his forecast to doubling every two years. This trend largely held until the early 2010s when improvements in transistor density began slowing.

AI Algorithms Require Advanced Chips for Economic and Efficiency Reasons

Leading-edge AI applications require state-of-the-art chips for both economic and efficiency reasons. According to CSET, training an AI algorithm can cost tens of millions of US dollars and take weeks to complete even when the most advanced AI chips are used.[155] By some estimates, the training time for AlphaGo—the first computer program to beat a world champion in the game "Go"—cost US$35 million and took between four and six weeks to complete.[156,157] Older chips consume more power and take far longer to perform the computations necessary for functions such as training and inference. The timely development and deployment of cost-effective leading-edge AI applications thus generally requires advanced or state-of-the-art AI chips.*

Power, Throughput, and Efficiency Are Key Metrics for Evaluating Chip Performance

The goal in the design of most chips is to maximize efficiency, which also pertains to chips designed for AI applications. *Efficiency* depends on throughput and power—it is equal to the throughput divided by the power, so the goal is to maximize the throughput while minimizing power consumption. *Power*—alternatively referred to as the thermal design power, or TDP—refers to the amount of heat the chip dissipates at maximum capacity (in watts). Power is a critical variable because high power consumption can limit the types of systems in which the chip can be implemented. *Throughput* refers to the amount of operations per unit time a processor can perform—this is usually counted in TOPS with today's most advanced capabilities.

In addition to power, throughput, and efficiency, other key parameters for evaluating the performance of an AI chip include data precision; memory size, clock, and bandwidth; and the chip type. *Data precision* refers to the method used to represent data for processor storage and calculation; a more optimized data format will enable the chip to perform a higher number of operations per second. However, there is a tradeoff between precision and efficiency; the smaller the size of the data units—the methods can range from integers (INT) to floating point (FP)—the greater the throughput. The *memory size*, *clock*, and *bandwidth* dictate how much information the chips can work with and how they communicate. These parameters are critical for evaluating chip performance because they can indicate processing speed. The chip type, meanwhile, can help dictate the types of applications for which the chip is suited, as discussed in the following section.

Different Types of AI Chips Are Suited for Distinct Applications

The main classes of AI chips include GPUs, FPGAs, ASICs, and SoCs.[158] As noted above, different types of chips are suited for distinct applications.

- *GPUs* are processors that use highly parallel architectures to perform simultaneous operations. In contrast to CPUs, which execute instructions serially on a single core, GPUs have concurrent threads running on multiple cores to speed the process considerably.[159] GPUs are primarily used for training AI algorithms, and—to a more limited extent—for inference applications.[160,161,162]

* As CSET notes, the need for substantial computational horsepower could change with the advancement of AI techniques such as few shot learning that do not rely on highly trained models.

- *FPGAs* are integrated circuits containing a preset number of logic gates that can be switched on or off by the customer; as such, they offer comparatively more flexibility in design (and, by extension, efficiency) than CPUs and GPUs as new adjustments are needed. Because of their greater efficiency, FPGAs are an attractive option for inference functions.[*,163]
- *ASICs* are highly efficient (even more so than FPGAs) and optimized integrated circuits that are programmed and constructed for a specific algorithm. The drawback of ASICs relative to other types of chips is that once the ASIC is constructed, it cannot be adjusted for future changes in the algorithm. ASICs thus maximize efficiency at the expense of flexibility. ASICs can be designed for training or inference functions.[†,164,165]
- An *SoC*—alternatively called a "machine on a chip (MoC)" or a "network on a chip (NoC)"—is an integrated circuit where all of the necessary processors, controllers, and memory to allow for the full function of the AI chip are on a single chip; this design feature is critical for reducing the accelerator's latency. In many cases, SoCs combine traditional computing architectures with various hardware and software acceleration schemes to optimize performance. As such, an SoC may incorporate a CPU, GPU, and memory all on a single chip.

[*] While training applications almost always benefit from data parallelism, inference functions do not. Inference may, for example, be performed on a single piece of data at a time. For a more detailed discussion of parallelism for training versus inference, see Khan and Mann's 2020 article (reference 163).

[†] As noted by Nanalyze—a research firm focused on disruptive technologies—in a 2020 article (reference 165), custom-built ASICs for AI applications often go by other names. Examples include TPUs, NPUs, and IPUs.

References

1. Website | New America | Full Translation: China's 'New Generation Artificial Intelligence Development Plan' (2017) | https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017 | Accessed 31 March 2020 | G. Webster, R. Creemers, P. Triolo, and E. Kania, authors.
2. Report | Deciphering China's AI Dream | March 2018 | p. 23 | J. Ding, Future of Humanity Institute, University of Oxford, author.
3. Report | Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security | February 2019 | p. 5 | G.C. Allen (Center for a New American Security), author.
4. Report | Chinese Military Innovation in Artificial Intelligence: Testimony before the U.S.-China Economic and Security Review Commission Hearing on Trade, Technology, and Military-Civil Fusion | Available at https://www.uscc.gov/sites/default/files/%207%20Hearing_Panel%201_Elsa%20Kania_Chinese%20Military%20Innovation%20in%20Artificial%20Intelligence_0.pdf | 7 June 2019 | Accessed 5 June 2020 | E. Kania (Center for Security and Emerging Technology, Georgetown University), author.
5. Website | 199: Intelligentization and a Chinese Vision of Future War | https://madsciblog.tradoc.army.mil/199-intelligentization-and-a-chinese-vision-of-future-war/ | 19 December 2019 | Accessed 5 June 2020 | Mad Scientist Laboratory Blog.
6. Report | Learning the Superior Techniques of the Barbarians: China's Pursuit of Semiconductor Independence | January 2019 | p. 8 | J.A. Lewis (CSIS), author.
7. Website | MIT Technology Review | China has never had a real chip industry. Making AI Chips could change that. | https://www.technologyreview.com/2018/ 12/14/138260/china-has-never-had-a-real-chip-industry-making-ai-chips-could-change-that/ | 14 December 2018 | Accessed 9 April 2020 | W. Knight, author.
8. Website | MIT Technology Review | China has never had a real chip industry. Making AI Chips could change that. | https://www.technologyreview.com/2018/ 12/14/138260/china-has-never-had-a-real-chip-industry-making-ai-chips-could-change-that/ | 14 December 2018 | Accessed 9 April 2020 | W. Knight, author.
9. Website | MacroPolo | AI Chips | https://macropolo.org/ digital-projects/supply-chain/ai-chips/inside-an-ai-chip/ | Accessed 14 April 2020 | MacroPolo's website is subtitled "Decoding China's Economic Arrival."
10. Article | *PC Gamer* | Chipmaker TSMC is spending billions of dollars to meet 7nm and 5nm demand | Available at https://www.pcgamer.com/chipmaker-tsmc-is-spending-billions-of-dollars-to-meet-7nm-and-5nm-demand/ | 24 October 2019 | Accessed 9 April 2020 | P. Lilly, author.
11. Article | *VentureBeat* | Why the $10 billion chip factory club just got smaller | Available at https://venturebeat.com/2018/08/28/why-the-10-billion-chip-factory-club-just-got-smaller/ | 28 August 2018 | Accessed 9 April 2020 | D. Takahashi, author.
12. Article | *Engadget* | Major AMD chip supplier will no longer make next-gen chips | https://www.engadget.com/2018-08-28-global-foundries-stops-7-nanometer-chip-production.html | 28 August 2018 | Accessed 9 April 2020 | S. Dent, author.
13. Website | MacroPolo | AI Chips | https://macropolo.org/ digital-projects/supply-chain/ai-chips/inside-an-ai-chip/ | Accessed 14 April 2020 | MacroPolo's website is subtitled "Decoding China's Economic Arrival."
14. Article | *Bloomberg* | Behind Samsung's $116 Billion Bid for Chip Supremacy | Available at https://www.bloomberg.com/news/articles/2019-12-23/behind-samsung-s-116-billion-bid-for-chip-supremacy | 22 December 2019 | Accessed 9 April 2020 | Sohee Kim, author.

15. Article | *Bloomberg* | Behind Samsung's $116 Billion Bid for Chip Supremacy | Available at https://www.bloomberg.com/news/articles/2019-12-23/behind-samsung-s-116-billion-bid-for-chip-supremacy | 22 December 2019 | Accessed 9 April 2020 | Sohee Kim, author.
16. Article | *Quartz* | Why the semiconductor is suddenly at the heart of US-China tech tensions | Available at https://qz.com/1335801/us-china-tech-why-the-semiconductor-is-suddenly-at-the-heart-of-us-china-tensions/ | 24 July 2018 | Accessed 14 April 2020 | J. Horwitz, author.
17. Website | Real Money | Huawei's Work on Alternatives to U.S. Tech -- And the Challenges They Face | Available at https://realmoney.thestreet.com/investing/technology/huawei-s-work-on-alternatives-to-u-s-tech-and-the-challenges-they-face-15072796 | Accessed 27 April 2020 | E. Jhonsa, author.
18. Website | Wired | China Challenges Nvidia's Hold on Artificial Intelligence Chips | Available at https://www.wired.com/story/china-challenges-nvidias-hold-on-artificial-intelligence-chips/ | 20 November 2017 | Accessed 6 May 2020 | T. Simonite, author.
19. Article | *Technode* | SILICON – Why Chinese EDA Tools Lag Behind | Available at https://technode.com/2019/11/13/silicon-why-chinese-eda-tools-lag-behind/ | 13 November 2019 | Accessed 27 April 2020 | S. Randall, author.
20. Report | Learning the Superior Techniques of the Barbarians: China's Pursuit of Semiconductor Independence | January 2019 | p. 8 | J.A. Lewis (CSIS), author.
21. Article | *Quartz* | Why the semiconductor is suddenly at the heart of US-China tech tensions | Available at https://qz.com/1335801/us-china-tech-why-the-semiconductor-is-suddenly-at-the-heart-of-us-china-tensions/ | 24 July 2018 | Accessed 14 April 2020 | J. Horwitz, author.
22. Website | Kirin Product List | http://www.hisilicon.com/en/Products/ProductList/Kirin | Accessed 19 June 2020 | HiSilicon product evaluation website.
23. Website | Synced | Huawei 7nm Kirin 810 Beats Snapdragon 855 and Kirin 980 on AI Benchmark Test | 21 June 2019 | https://syncedreview.com/2019/06/21/huawei-7nm-kirin-810-beats-snapdragon-855-and-kirin-980-on-ai-benchmark-test/ | Accessed 19 June 2020 | Commercial website.
24. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 27 | S.M. Khan and A. Mann (CSET), authors.
25. Website | Bald Engineering Blog | Chinese AMEC 5nm plasma etching tools verified by TSMC | http://www.blog.baldengineering.com/2018/12/chinese-amec-5nm-plasma-etching-tools.html | 18 December 2018 | Accessed 22 April 2020 | Commercial website.
26. Website | HSMC Products and Services | http://en.hsmc.com/intro/6.html | Accessed 22 April 2020 | Commercial website.
27. Website | 产品技术 | AMEC | https://www.amec-inc.com/products.html#Etch | Accessed 22 April 2020 | Commercial website.
28. Website | Bald Engineering Blog | Chinese AMEC 5nm plasma etching tools verified by TSMC | http://www.blog.baldengineering.com/2018/12/chinese-amec-5nm-plasma-etching-tools.html | 18 December 2018 | Accessed 22 April 2020 | Commercial website.
29. Article | *Reuters* | Trump administration pressed Dutch hard to cancel China chip-equipment sale: sources | Available at https://www.reuters.com/article/us-asml-holding-usa-china-insight/trump-administration-pressed-dutch-hard-to-cancel-china-chip-equipment-sale-sources-idUSKBN1Z50HN | 5 January 2020 | Accessed 22 April 2020 | A. Alper, T. Sterling, and S. Nellis, authors.
30. Website | About us | HSMC | http://en.hsmc.com/intro/3.html | Accessed 22 April 2020 | Commercial website.
31. Website | AMEC | Primo D-RIE | https://www.amec-inc.com/productDetails.html?id=5d1d7361f5ec603023758195 | Accessed 24 June 2020 | Commercial website.

32. Website | About us | HSMC | http://en.hsmc.com/intro/3.html | Accessed 24 June 2020 | Commercial website.
33. Website | HSMC Products and Services | http://en.hsmc.com/intro/6.html | Accessed 24 June 2020 | Commercial website.
34. Website | Ushio | Products | "Litho-Patterning" | EUV light source | https://www.ushio.co.jp/en/products/keyword_details.html?func_tag=1101&key_label=Litho-Patterning | Accessed 20 July 2020 | Commercial website.
35. Website | Canon | Semiconductor Lithography Equipment | https://global.canon/en/product/indtech/semicon/index.html | Accessed 24 June 2020 | Commercial website.
36. Website | Canon | FPA-6300ES6a | https://sg.canon/en/business/fpa-6300es6a/product | Accessed 24 June 2020 | Commercial website.
37. Website | Nikon | Products archive | https://www.nikon.com/products/semi/lineup/archives/index.htm | Accessed 24 June 2020 | Commercial website.
38. Website | Accretech | Semiconductor Production Process | https://www.accretech.jp/english/product/semicon/about.html | Accessed 24 June 2020 | Japanese English-language website.
39. Website | TEL | Tactras | https://www.tel.co.jp/product/tactras.html | Accessed 24 June 2020 | Japanese English-language website.
40. Website | ASML | EUV Lithography Systems | https://www.asml.com/en/products/euv-lithography-systems | Accessed 24 June 2020 | Commercial website.
41. Website | ASML | Twinscan NXE:3400C | https://www.asml.com/en/products/euv-lithography-systems/twinscan-nxe3400c | Accessed 24 June 2020 | Commercial website.
42. Website | KLA | Metrology | https://www.kla-tencor.com/products/chip-manufacturing/metrology | Accessed 24 June 2020 | Commercial website.
43. Website | Lam Research | Our Products | https://www.lamresearch.com/products/our-products/ | Accessed 24 June 2020 | Commercial website.
44. Website | Lam Research | Sola Product Family | https://www.lamresearch.com/product/sola-product-family/ | Accessed 24 June 2020 | Commercial website.
45. Website | Applied Materials | Semiconductor Products | http://www.appliedmaterials.com/semiconductor/products | Accessed 24 June 2020 | Commercial website.
46. Website | Applied Materials | Centura Tetra EUV Advanced Reticle Etch | http://www.appliedmaterials.com/products/centura-tetra-euv-advanced-reticle-etch | Accessed 24 June 2020 | Commercial website.
47. Article | *Bloomberg* | Behind Samsung's $116 Billion Bid for Chip Supremacy | Available at https://www.bloomberg.com/news/articles/2019-12-23/behind-samsung-s-116-billion-bid-for-chip-supremacy | 22 December 2019 | Accessed 9 April 2020 | Sohee Kim, author.
48. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 30 | S.M. Khan and A. Mann (CSET), authors.
49. Article | *Wall Street Journal* | U.S. Sets Export Controls on China's Top Chip Maker | Available at https://www.wsj.com/articles/u-s-sets-export-controls-on-chinas-top-chip-maker-11601118353 | 28 September 2020 | Accessed 29 September 2020 | Dan Strumpf, author.
50. Article | *The New York Times* | U.S. Places Restrictions on China's Leading Chip Maker | Available at https://www.nytimes.com/2020/09/26/technology/trump-china-smic-blacklist.html | 26 September 2020 | Accessed 29 September 2020 | Ana Swanson and Raymond Zhong, authors.
51. (U) Article | *Barron's* | China Chip Giant SMIC Shares Sink On US Export Controls | Available at https://www.barrons.com/amp/news/china-chip-giant-smic-shares-dive-on-us-export-controls-01601281805 | 28 September 2020 | Accessed 29 September 2020 | Jerome Taylor, author.

52. Website | UMC | About UMC | http://www.umc.com/japanese/about/index.asp | Accessed 28 June 2020 | Commercial website.

53. Website | *Anand Tech* | SMIC To Start 14nm Mass Production in H1 2019 | https://www.anandtech.com/show/13941/smics-14-nm-mass-production-in-1h-2019 | 8 February 2019 | Accessed 28 June 2020 | A. Shilov, author.

54. Website | Global Foundries | 12LP | https://www.globalfoundries.com/sites/default/files/ product-briefs/pb-12lp-11-web.pdf | Accessed 28 June 2020 | Commercial website.

55. Website | Intel | Intel 14 nm Technology | https://www.intel.com/content/www/us/en/silicon-innovations/intel-14nm-technology.html | Accessed 28 June 2020 | Commercial website.

56. Website | *Anand Tech* | Samsung Starts Mass Production at V1: A Dedicated EUV Fab for 7nm, 6nm, 5nm, 4nm, 3nm Nodes | https://www.anandtech.com/show/15538/samsung-starts-mass-production-at-v1-a-dedicated-euv-fab-for-7nm-6nm-5nm-4nm-3nm-nodes | 20 February 2020 | Accessed 28 June 2020 | A. Shilov, author.

57. Website | *Tom's Hardware* | Intel To Produce 7nm Node At Fab 42, Announces $7 Billion Investment To Finish Construction | https://www.tomshardware.com/news/intel-fab-42-7mn-chandler,33619.html | 8 February 2017 | Accessed 28 June 2020 | P. Alcorn, author.

58. Article | *Nikkei Asian Review* | TSMC to spend $20bn on 3-nanometer chips | Available at https://asia.nikkei.com/Business/TSMC-to-spend-20bn-on-3-nanometer-chips | 14 December 2017 | Accessed 28 June 2020 | Cheng Ting-Fang, author.

59. Website | TSMC | GIGAFAB Facilities | https://www.tsmc.com/english/dedicatedFoundry/ manufacturing/gigafab.htm | Accessed 28 June 2020 | Commercial website.

60. Article | *Bloomberg* | Battle In AI Chips Heats Up As Nvidia Faces A Brigade Of Rivals | https://www.investors.com/news/technology/ai-chips-nvidia-artificial-intelligence-stocks/ | 30 December 2019 | Accessed 24 April 2020 | R. Krause, author.

61. Website | Huawei | Huawei launches Ascend 910, the world's most powerful AI processor, and MindSpore, an all-scenario AI computing framework | https://www.huawei.com/us/news/2019/8/ huawei-ascend-910-most-powerful-ai-processor | 23 August 2019 | Accessed 28 June 2020 | Commercial website.

62. Website | Huawei | Ascend 910 AI Processor | https://e.huawei.com/us/products/cloud-computing-dc/atlas/ascend-910 | Accessed 28 June 2020 | Accessed 28 June 2020 | Commercial website.

63. Website | Huawei | Driving AI to new horizons with an all-scenario native, full-stack solution | https://www.huawei.com/en/publications/communicate/86/driving-ai-to-new-horizons | Accessed 28 June 2020 | Dang Wenshuan, author.

64. Website | Huawei | The world's most powerful AI processor | https://forum.huawei.com/ enterprise/en/the-world-s-most-powerful-ai-processor/thread/568113-895 | 18 September 2019 | Accessed 28 June 2020 | Commercial website.

65. Website | Enflame Technology Announces CloudBlazer with DTU Chip on GLOBALFOUNDRIES 12LP FinFET Platform for Data Center Training | https://www. globenewswire.com/news-release/2019/12/12/1959813/0/en/Enflame-Technology-Announces-CloudBlazer-with-DTU-Chip-on-GLOBALFOUNDRIES-12LP-FinFET-Platform-for-Data-Center-Training.html | 12 December 2019 | Accessed 28 June 2020 | Commercial website.

66. Website | Enflame Tech | Cloudblazer | https://www.enflame-tech.com/product-technologies/T11 | Accessed 28 June 2020 | Commercial website.

67. Website | Slideshare.net | Cloud Deep Learning Chips – Training & Inference | https://www. slideshare.net/ssuser479fa3/cloud-deep-learning-chips-training-inference | 28 December 2019 | Accessed 28 June 2020 | Commercial website.

68. Website | Nvidia | Nvidia Turing GPU Architecture | https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf | Accessed 28 June 2020 | Commercial website.
69. Website | Nvidia Developer | NVIDIA AI Inference Performance Milestones: Delivering Leading Throughput, Latency and Efficiency | 12 November 2018 | https://news.developer.nvidia.com/nvidia-ai-inference-performance-milestones-delivering-leading-throughput-latency-and-efficiency/ | Accessed 28 June 2020 | Commercial website.
70. Website | Tech Power Up | Nvidia Tesla T4 | https://www.techpowerup.com/gpu-specs/tesla-t4.c3316#:~:text=The%20Tesla%20T4%20is%20a,card%20supports%20DirectX%2012%20Ultimate. | Accessed 28 June 2020 | Commercial website.
71. Website | Nvidia | Nvidia Tesla V100 GPU Architecture | https://images.nvidia.com/content/ volta-architecture/pdf/volta-architecture-whitepaper.pdf | August 2017 | Accessed 28 June 2020 | Commercial website.
72. Website | Nvidia Developer | NVIDIA AI Inference Performance Milestones: Delivering Leading Throughput, Latency and Efficiency | 12 November 2018 | https://news.developer.nvidia.com/nvidia-ai-inference-performance-milestones-delivering-leading-throughput-latency-and-efficiency/ | Accessed 28 June 2020 | Commercial website.
73. Website | Habana | Habana Goya – Inference Card | https://habana.ai/wp-content/uploads/ 2019/06/Goya-Datasheet-HL-10x.pdf | Accessed 29 June 2020 | Commercial website.
74. Website | Wikichip | A Look At The Habana Inference And Training Neural Processors | https://fuse.wikichip.org/news/3159/a-look-at-the-habana-inference-and-training-neural-processors/ | 15 December 2019 | Accessed 29 June 2020 | Public website.
75. Report | Habana | Gaudi Training Platform White Paper | Available at https://habana.ai/wp-content/uploads/2019/06/Habana-Gaudi-Training-Platform-whitepaper.pdf | June 2019, Ver 1.0 | Accessed 29 June 2020 | Online report from commercial website.
76. Website | Wikichip | A Look At The Habana Inference And Training Neural Processors | https://fuse.wikichip.org/news/3159/a-look-at-the-habana-inference-and-training-neural-processors/ | 15 December 2019 | Accessed 29 June 2020 | Public website.
77. Report | Habana | Hot Chips 2019 | https://www.hotchips.org/hc31/HC31_1.14_HabanaLabs. Eitan_Medina.v9.pdf | August 2019 | Accessed 29 June 2020 | Online report from commercial website.
78. Website | Habana | GAUDI AI Training | https://habana.ai/training/ | Accessed 29 June 2020 | Commercial website.
79. Website | Huawei | Kirin 980 | https://consumer.huawei.com/en/campaign/kirin980/ | Accessed 28 June 2020 | Commercial website.
80. Website | Huawei | Huawei P30 Pro | https://consumer.huawei.com/en/phones/p30-pro/specs/ | Accessed 29 June 2020 | Commercial website.
81. Website | Huawei | Huawei Kirin 990 Series – Rethink Evolution | https://consumer.huawei.com/ en/campaign/kirin-990-series/ | Accessed 29 June 2020 | Commercial website.
82. Website | Huawei | Huawei Mate 30 Pro 5G | https://consumer.huawei.com/en/phones/mate30-pro-5g/specs/ | Accessed 29 June 2020 | Commercial website.
83. Website | Qualcomm | Snapdragon 865 5G Mobile Platform | https://www.qualcomm.com/ products/snapdragon-865-5g-mobile-platform | Accessed 28 June 2020 | Commercial website.
84. Website | Nanoreview.net | Exynos 990 vs Snapdragon 865 | https://nanoreview.net/en/soc-compare/samsung-exynos-990-vs-qualcomm-snapdragon-865 | Accessed 29 June 2020 | Commercial website.

85. Website | WCCF Tech | Samsung Galaxy Tab S7 Will Have Snapdragon 865 and 8GB RAM | https://wccftech.com/samsung-galaxy-tab-s7-will-have-snapdragon-865-and-8gb-ram/ | 25 June [year unspecified, likely 2020] | Accessed 29 June 2020 | F. Shahid, author.
86. Website | Apple Newsroom | iPhone Xs and iPhone Xs Max bring the best and biggest displays to iPhone | 12 September 2018 | https://www.apple.com/newsroom/2018/09/iphone-xs-and-iphone-xs-max-bring-the-best-and-biggest-displays-to-iphone/ | Accessed 29 June 2020 | Commercial website.
87. Website | Apple Insider | iPhone XS has 4GB of RAM, 2.49 GHz A12 chip according to benchmarks | https://appleinsider.com/articles/18/09/13/iphone-xs-has-4gb-of-ram-249-ghz-a12-chip-according-to-benchmarks | Accessed 29 June 2020 | R. Fingas, author.
88. Website | Anand Tech | Apple Announces New iPhone 11, iPhone 11 Pro, & iPhone 11 Pro Max | https://www.anandtech.com/show/14859/apple-announces-new-iphone-11-iphone-11-pro-iphone-11-pro-max | 10 September 2019 | Accessed 28 June 2020 | A. Frumusanu, author.
89. Website | Techspot | Apple's A13 SoC will be the first chip manufactured using new 7nm 'N7 Pro' process | https://www.techspot.com/news/79657-apple-chips-first-use-new-7nm-n7-pro.html | 15 April 2019 | Accessed 29 June 2020 | R. Thubron, author.
90. Website | Apple | Apple Special Event – September 10, 2019 | https://www.apple.com/apple-events/september-2019/ | Accessed 29 June 2020 | Commercial website.
91. Website | Wired | An Exclusive Look Inside Apple's A13 Bionic Chip | https://www.wired.com/story/apple-a13-bionic-chip-iphone/ | 19 September 2019 | Accessed 29 June 2020 | Commercial website.
92. Website | Anand Tech | The Apple iPhone 11, 11 Pro & 11 Pro Max Review: Performance, Battery, & Camera Elevated | https://www.anandtech.com/show/14892/the-apple-iphone-11-pro-and-max-review/2 | 16 October 2019 | Accessed 29 June 2020 | A. Frumusanu, author.
93. Website | Anand Tech | Apple Announces New iPhone 11, iPhone 11 Pro, & iPhone 11 Pro Max | https://www.anandtech.com/show/14859/apple-announces-new-iphone-11-iphone-11-pro-iphone-11-pro-max | 10 September 2019 | Accessed 29 June 2020 | A. Frumusanu, author.
94. Website | Techspot | Apple's A13 SoC will be the first chip manufactured using new 7nm 'N7 Pro' process | https://www.techspot.com/news/79657-apple-chips-first-use-new-7nm-n7-pro.html | 15 April 2019 | Accessed 29 June 2020 | R. Thubron, author.
95. Website | Google Support | Pixel phone hardware tech specs | https://support.google.com/pixelphone/answer/7158570?hl=en | Accessed 29 June 2020 | Commercial website.
96. Website | Google | Cloud Tensor Processing Units (TPUs) | https://cloud.google.com/ tpu/docs/tpus | Accessed 29 June 2020 | Commercial website.
97. Website | Google | Release Notes | https://cloud.google.com/tpu/docs/release-notes | Accessed 29 June 2020 | Commercial website.
98. Website | Google | Cloud TPU | https://cloud.google.com/tpu | Accessed 29 June 2020 | Commercial website.
99. Website | Amazon Blog | Amazon EC2 Update – Inf1 Instances with AWS Inferentia Chips for High Performance Cost-Effective Inferencing | https://aws.amazon.com/blogs/aws/amazon-ec2-update-inf1-instances-with-aws-inferentia-chips-for-high-performance-cost-effective-inferencing/ | 3 December 2019 | Accessed 29 June 2020 | J. Barr, author.
100. Website | Amazon | AWS Inferentia | https://aws.amazon.com/machine-learning/inferentia/?nc2=h_ql_prod_ml_inf | Accessed 29 June 2020 | Commercial website.
101. Website | Microsoft | Windows Virtual Machine Pricing | https://azure.microsoft.com/en-us/pricing/details/virtual-machines/windows/ | Accessed 28 June 2020 | Commercial website.
102. Website | Microsoft | Inside the Microsoft FPGA-based configurable cloud | https://azure.microsoft.com/en-us/resources/videos/build-2017-inside-the-microsoft-fpga-based-configurable-cloud/ | Accessed 29 June 2020 | Commercial website.

103. Website | cnTechPost | Alibaba chipmaking arm expected to become a major TSMC customer | https://cntechpost.com/2020/05/07/alibaba-chipmaking-arm-expected-to-become-a-major-tsmc-customer/ | 7 May 2020 | Accessed 29 June 2020 | Commercial website.
104. Website | Alibaba Cloud | Announcing Hanguang 800: Alibaba's First AI-Inference Chip | Alibaba Cloud | https://www.alibabacloud.com/blog/announcing-hanguang-800-alibabas-first-ai-inference-chip_595482 | 28 October 2019 | Accessed 29 June 2020 | Commercial website.
105. Website | Samsung Newsroom | Baidu and Samsung Electronics Ready for Production of Leading-Edge AI Chip for Early Next Year | https://news.samsung.com/global/baidu-and-samsung-electronics-ready-for-production-of-leading-edge-ai-chip-for-early-next-year | 18 December 2019 | Accessed 29 June 2020 | Commercial website.
106. Website | Cambricon | 产品技术 | http://www.cambricon.com/index.php?m=content&c =index&a=lists&catid=15 | Accessed 29 June 2020 | Commercial website.
107. Website | J Q News | A New Generation of Cloud AI Chip Siyuan 270 in Cambrian Period | https://www.jqknews.com/news/208967-A_New_Generation_of_Cloud_AI_Chip_Siyuan_270_in _Cambrian_Period.html | 21 June 2019 | Accessed 29 June 2020 | Commercial website.
108. Article | *Nature* | Towards artificial general intelligence with hybrid Tianjic chip architecture | Vol. 572 | August 2019 | Jing Pei et al., authors.
109. Article | *Nature* | Towards artificial general intelligence with hybrid Tianjic chip architecture | Vol. 572 | August 2019 | Jing Pei et al., authors.
110. Website | Medium | Nature Cover Story – Chinese Team's 'Tianjic Chip' Bridges Machine Learning and Neuroscience in Pursuit of AGI | https://medium.com/syncedreview/nature-cover-story-chinese-teams-tianjic-chip-bridges-machine-learning-and-neuroscience-in-f1c3e8a03113 | 31 July 2019 | Accessed 6 May 2020 | Tony Peng and Fangyu Cai, authors.
111. Article | *Nature* | Towards artificial general intelligence with hybrid Tianjic chip architecture | Vol. 572 | August 2019 | Jing Pei et al., authors.
112. Website | MotivNt | Нейрочип «Алтай» | https://motivnt.ru/neurochip-altai/ | Accessed 29 June 2020 | Russian-language website.
113. Website | Loihi – a brief introduction | Intel | https://niceworkshop.org/wp-content/uploads/ 2018/05/Mike-Davies-NICE-Loihi-Intro-Talk-2018.pdf | Accessed 29 June 2020 | M. Davies, author.
114. Website | Extreme Tech | Intel's Neuromorphic Loihi Processor Scales to 8M Neurons, 64 Cores | https://www.extremetech.com/computing/295043-intels-neuromorphic-loihi-processor-scales-to-8m-neurons-64-cores | 16 July 2019 | Accessed 29 June 2020 | J. Hruska, author.
115. Website | IBM | The brain's architecture, efficiency…on a chip | https://www.ibm.com/blogs/ research/2016/12/the-brains-architecture-efficiency-on-a-chip/ | 19 December 2016 | Accessed 29 June 2020 | Commercial website.
116. Website | Introducing a Brain-inspired Computer | IBM Research | http://www.research.ibm.com/ articles/brain-chip.shtml | Accessed 29 June 2020 | D.S. Modha, author.
117. Article | *Design, Automation & Test in Europe Conference & Exhibition, 2017* | Understanding the design of IBM neurosynaptic system and its tradeoffs: A user perspective | IEEE conference, 27–31 March 2017 | pp. 139–144 | doi: 10.23919/DATE.2017.7926972 | H. Cheng et al., authors.
118. Website | IBM | Deep learning inference possible in embedded systems thanks to TrueNorth | https://www.ibm.com/blogs/research/2016/09/deep-learning-possible-embedded-systems-thanks-truenorth/ | 21 September 2016 | Accessed 29 June 2020 | Commercial website.
119. Article | *Science* | A million spiking-neuron integrated circuit with a scalable communication network and interface | Vol. 345, Iss. 6197 | 8 August 2014 | pp. 668–673 | P.A. Merolla et al., authors.

120. Website | Towards Data Science | Neuromorphic Hardware: Trying to Put Brain Into Chips | https://towardsdatascience.com/neuromorphic-hardware-trying-to-put-brain-into-chips-222132f7e4de | 30 June 2019 | Accessed 30 April 2020 | S. Gupta, authors.
121. Website | Intel | Beyond AI: New algorithmic approaches emulate the human brain's interactions with the world | https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html | Accessed 1 May 2020 | Commercial website.
122. Website | Towards Data Science | Neuromorphic Hardware: Trying to Put Brain Into Chips | https://towardsdatascience.com/neuromorphic-hardware-trying-to-put-brain-into-chips-222132f7e4de | 30 June 2019 | Accessed 30 April 2020 | S. Gupta, authors.
123. Article | *Science Magazine* | A million spiking-neuron integrated circuit with a scalable communication network and interface | Vol. 345, Iss. 6197 | 8 August 2014 | P.A. Merolla et al., authors.
124. Website | EE Times Asia | Neuromorphic Computing vs AI Chips: Compete or Complement? | https://www.eetasia.com/neuromorphic-computing-vs-ai-chips-compete-or-complement/ | 15 April 2020 | Accessed 20 April 2020 | S. Ward-Foxton, author.
125. Website | Intel | Beyond AI: New algorithmic approaches emulate the human brain's interactions with the world | https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html | Accessed 1 May 2020 | Commercial website.
126. Article | arXiv:2001.10362 [eess.SP] | The Final Frontier: Deep Learning in Space | 3 February 2020 | V. Kothari, E. Liberis, and N.D. Lane, authors.
127. Article | arXiv:2001.10362 [eess.SP] | The Final Frontier: Deep Learning in Space | 3 February 2020 | V. Kothari, E. Liberis, and N.D. Lane, authors.
128. Article | arXiv:2001.10362 [eess.SP] | The Final Frontier: Deep Learning in Space | 3 February 2020 | V. Kothari, E. Liberis, and N.D. Lane, authors.
129. Website | Aerospace | Delivering Artificial Intelligence to Space | https://aerospace.org/article/space-cloud-delivering-artificial-intelligence-space | 7 April 2019 | Accessed 4 May 2020 | Public website.
130. Website | Aerospace | Delivering Artificial Intelligence to Space | https://aerospace.org/article/space-cloud-delivering-artificial-intelligence-space | 7 April 2019 | Accessed 4 May 2020 | Public website.
131. Website | Aerospace | Delivering Artificial Intelligence to Space | https://aerospace.org/article/space-cloud-delivering-artificial-intelligence-space | 7 April 2019 | Accessed 4 May 2020 | Public website.
132. Website | Intel | Intel Vision Accelerator Design With Intel Movidius Vision Processing Unit (VPU) | https://software.intel.com/en-us/iot/hardware/vision-accelerator-movidius-vpu | Accessed 4 May 2020 | Commercial website.
133. Website | Aerospace | Delivering Artificial Intelligence to Space | https://aerospace.org/article/space-cloud-delivering-artificial-intelligence-space | 7 April 2019 | Accessed 4 May 2020 | Public website.
134. Website | Xilinx | Space-grade Virtex-5QV FPGA | https://www.xilinx.com/products/silicon-devices/fpga/virtex-5qv.html | Accessed 5 May 2020 | Commercial website.
135. Website | European Space Agency | ESA team blasts Intel's new AI chip with radiation at CERN | 29 November 2018 | https://www.esa.int/Enabling_support/Space_Engineering_Technology/ESA_team_blasts_Intel_s_new_AI_chip_with_radiation_at_CERN | Accessed 4 May 2020 | Public website.
136. Website | European Space Agency | First Earth observation satellite with AI ready for launch | http://www.esa.int/Applications/Observing_the_Earth/First_Earth_observation_satellite_with_AI_ready_for_launch | 9 December 2019 | Accessed 4 May 2020 | Public website.

137. Website | Vox | Intel is buying Movidius, a startup that makes vision trips for drones and virtual reality | https://www.vox.com/2016/9/6/12810246/intel-buying-movidius | 6 September 2016 | Accessed 5 May 2020 | I. Fried, author.
138. Website | Futurism.com | Intel's New Chip Helps Devices See and Interact with Their Environments in Real Time | https://futurism.com/intels-new-chip-helps-devices-see-and-interact-with-their-environments-in-real-time | 30 August 2020 | Accessed 4 May 2020 | K. Lant, author.
139. Website | European Space Agency | First Earth observation satellite with AI ready for launch | http://www.esa.int/Applications/Observing_the_Earth/First_Earth_observation_satellite_with_AI_ready_for_launch | 9 December 2019 | Accessed 4 May 2020 | Public website.
140. (U) Article | *Phys.org* | ESA selects follow-up AI Earth observatory satellite mission | Available at: https://phys.org/news/2020-09-esa-follow-up-ai-earth-observatory.html | 7 September 2020 | Accessed 29 September 2020 | European Space Agency, author.
141. Website | European Space Agency | https://www.esa.int/ESA_Multimedia/Images/2018/11/Myriad_2 | 28 November 2018 | Accessed 20 July 2020 | Public website.
142. Website | Bloomberg | The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies | Available at https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies | 4 October, 2018 | Accessed 29 September 2020 | Jordan Robertson and Michael Riley, authors.
143. Website | Wired | The US Fears Huawei Because It Knows How Tempting Backdoors Are | Available at https://www.wired.com/story/huawei-backdoors-us-crypto-ag/ | 11 February 2020 | Accessed 29 September 2020 | Lily Hay Newman, author.
144. Article | *Electronics Magazine* | Cramming More Components onto Integrated Circuits | 1965 | G. Moore, author.
145. Report | AI Chips: What They Are and Why They Matter | April 2020 | pp. 7–11 | S.M. Khan and A. Mann (CSET), authors.
146. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 10 | S.M. Khan and A. Mann (CSET), authors.
147. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 11 | S.M. Khan and A. Mann (CSET), authors.
148. Website | *arXiv* | Squeeze-and-Excitation Networks | https://arxiv.org/abs/1709.01507 | 5 September 2017 | Accessed 20 July 2020 | Jie Hu, Li Shen, and Gang Sun, authors.
149. Report | Microsoft Research Technical Report MSR-TR-2017-39 | *The Microsoft 2017 Conversational Speech Recognition System* | 2017 | Available at https://www.microsoft.com/en-us/research/wp-content/uploads/2017/08/ms_swbd17-2.pdf | W. Xiong et al., authors.
150. Article | *Nature* | Human-level control through deep reinforcement learning | Vol. 518 | 26 February 2015 | pp. 529–533 | V. Mnih et al., authors.
151. Website | *Nature* | Google AI beats top human players at strategy game *StarCraft II* | Available at https://www.nature.com/articles/d41586-019-03298-6 | 30 October 2019 | Accessed 15 June 2020 | D. Garisto, author.
152. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 34 | S.M. Khan and A. Mann (CSET), authors.
153. Website | MacroPolo | AI Chips | https://macropolo.org/ digital-projects/supply-chain/ai-chips/inside-an-ai-chip/ | Accessed 14 April 2020 | MacroPolo's website is subtitled "Decoding China's Economic Arrival."
154. Report | White Paper on AI Chip Technologies | 2018 | p. 7 | Zheng You and Shaojun Wei (Tsinghua University and Beijing Innovation Center for Future Chips), Eds.
155. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 6 | S.M. Khan and A. Mann (CSET), authors.

156. Website | *Wired* | DeepMind's Losses and the Future of Artificial Intelligence | https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/ | 14 August 2019 | Accessed 7 April 2020 | G. Marcus, author.

157. Website | Blog: *Deep Learning: Sky's the Limit?* | Part 2: AlphaGo under a Magnifying Glass (The historic match of deep learning AlphaGo vs. Lee Sedol) REVISED | 6 April 2016 | http://deeplearningskysthelimit.blogspot.com/2016/04/part-2-alphago-under-magnifying-glass.html | Accessed 7 April 2020 | Commercial website.

158. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 20 | S.M. Khan and A. Mann (CSET), authors.

159. Website | Towards Data Science | Neuromorphic Hardware: Trying to Put Brain Into Chips | https://towardsdatascience.com/neuromorphic-hardware-trying-to-put-brain-into-chips-222132f7e4de | 30 June 2019 | Accessed 30 April 2020 | S. Gupta, authors.

160. Report | AI Chips: What They Are and Why They Matter | April 2020 | p. 20 | S.M. Khan and A. Mann (CSET), authors.

161. Website | OpenAI Blog | AI and Compute | https://openai.com/blog/ai-and-compute/ | 16 May 2018 | Accessed 17 April 2020 | Commercial website.

162. Website | Nanalyze | How AI Chips Are Changing the Semiconductor Industry | https://www.nanalyze.com/2020/04/ai-chips-changing-semiconductor-industry/ | 20 April 2020 | Accessed 29 April 2020 | Commercial website.

163. Report | AI Chips: What They Are and Why They Matter | April 2020 | pp. 20–21 | S.M. Khan and A. Mann (CSET), authors.

164. Report | AI Chips: What They Are and Why They Matter | April 2020 | pp. 6, 20 | S.M. Khan and A. Mann (CSET), authors.

165. Website | Nanalyze | How AI Chips Are Changing the Semiconductor Industry | https://www.nanalyze.com/2020/04/ai-chips-changing-semiconductor-industry/ | 20 April 2020 | Accessed 29 April 2020 | Commercial website.