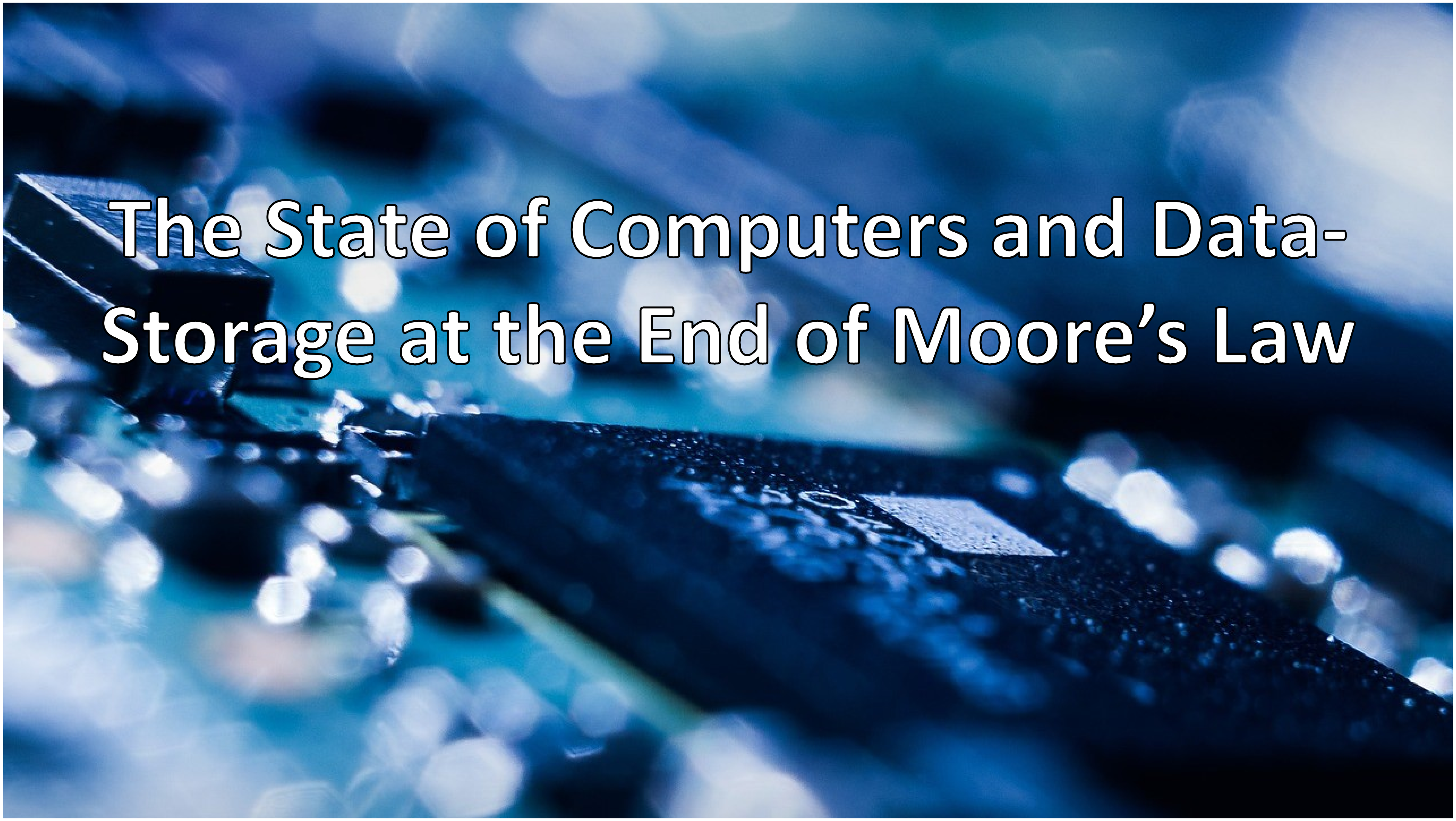
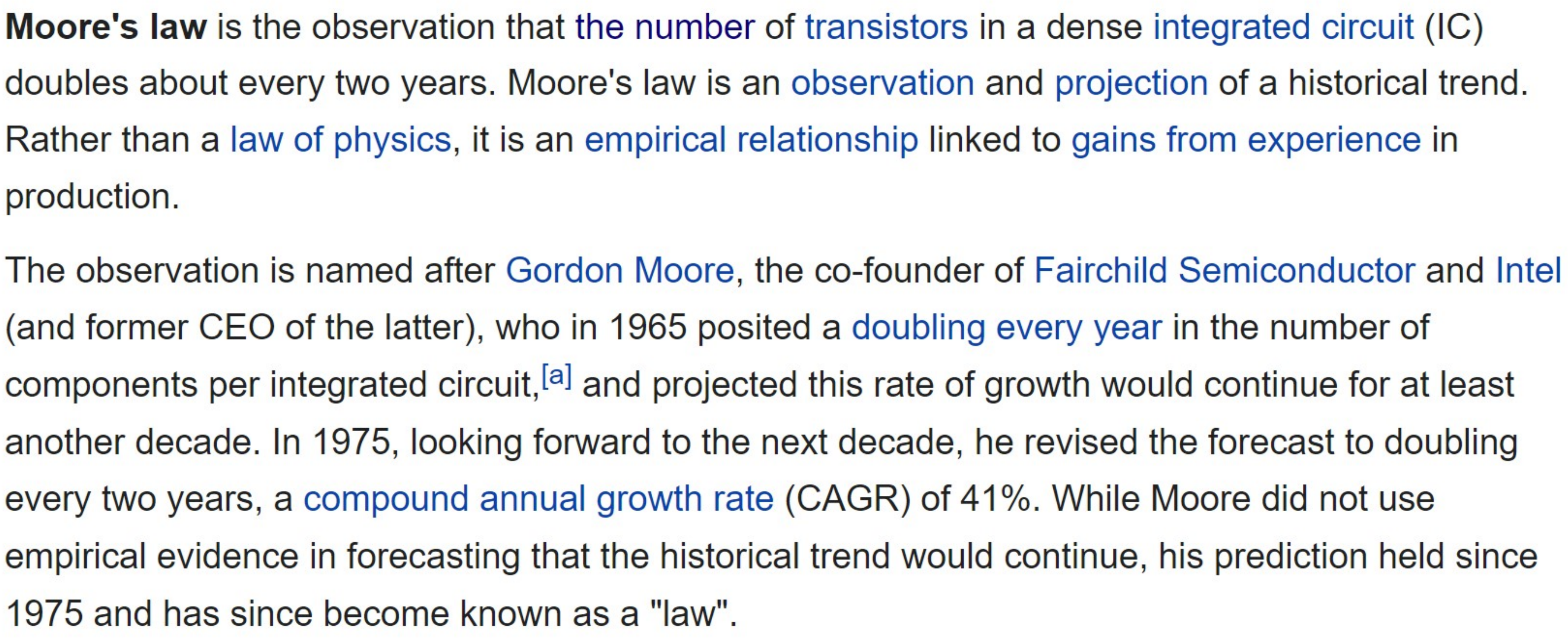


The State of Computers and Data-Storage at the End of Moore’s Law

First published 8. November, 2022
Latest revision 29. November, 2022

Definition



Moore's law is the observation that the number of transistors in a dense integrated circuit (IC) doubles about every two years. Moore's law is an observation and projection of a historical trend. Rather than a law of physics, it is an empirical relationship linked to gains from experience in production.

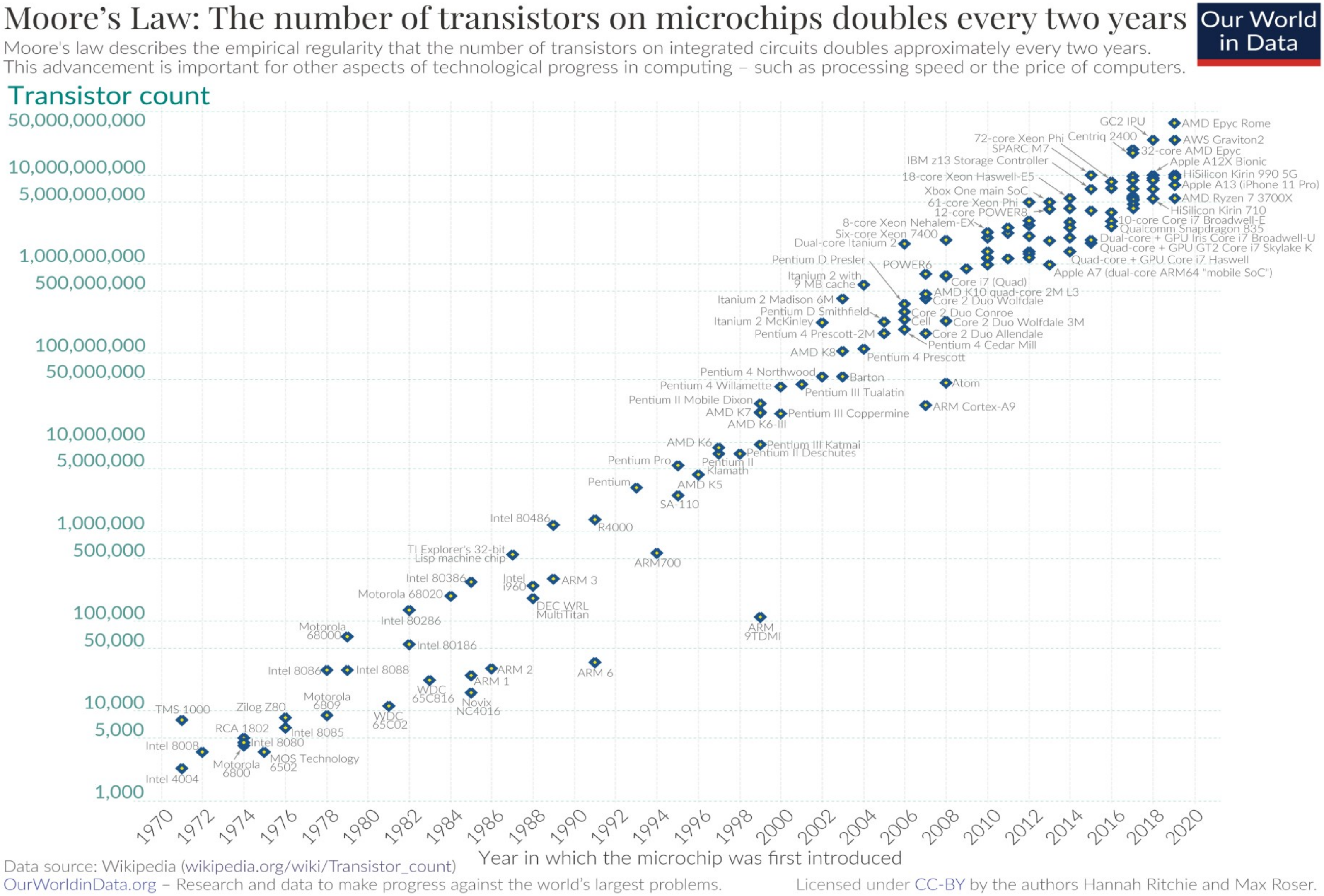
The observation is named after Gordon Moore, the co-founder of Fairchild Semiconductor and Intel (and former CEO of the latter), who in 1965 posited a doubling every year in the number of components per integrated circuit,[a] and projected this rate of growth would continue for at least another decade. In 1975, looking forward to the next decade, he revised the forecast to doubling every two years, a compound annual growth rate (CAGR) of 41%. While Moore did not use empirical evidence in forecasting that the historical trend would continue, his prediction held since 1975 and has since become known as a "law".

- **Source:** https://en.wikipedia.org/wiki/Moore%27s_law

Moore’s Law & The Future of Humanity

- ✓ **Integrated circuits (ICs) for computing and data storage** are massively important for the future of humanity
- ✓ **Such circuits enable all aspects of modern life**
- ✓ **Without ICs** the world would not have been able to progress its economies much beyond the level we had in the 1960s before the invention of these circuits
- ✓ **Nearly everything we currently produce** from physical goods to informational services make use of ICs directly or indirectly
- ✓ **Indeed, how far humanity can progress** its economies critically depends on further advances in making ICs
- ✓ **That is why Moore’s Law matters** and why it is worth investigating 1) **when** we can no longer squeeze more transistors into a circuit and 2) **how advanced** the ICs will be in the end

The Evidence



Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

OurWorldinData.org – Research and data to make progress against the world's largest problems.

- **Source:** https://en.wikipedia.org/wiki/Moore%27s_law#/media/File:Moore's_Law_Transistor_Count_1970-2020.png

Calculating growth rate from evidence

- ✓ **Graph shows state-of-the-art ICs in 1971** had about 5000 transistors
- ✓ **By 2019 a state-of-the-art IC** had about 40 billion transistors
- ✓ From 1971 to 2019 is 48 years
- ✓ **The compounded annual growth rate (CAGR)** can thus be calculated as
- ✓ **((40,000,000,000/5000)^(1/48))-1= 0.393**
- ✓ In percentage growth per year 39.3%
- ✓ **Growth in 2 years is calculated as**
- ✓ **1*(1+0.393)^2=1.94. If 10 years you get 27.425**
- ✓ **1.94 is practically 2 so Moore's law is still spot on with a doubling every second year in transistors per IC**

$$CAGR = \left(\frac{EV}{BV}\right)^{\frac{1}{n}} - 1$$

where:

$EV =$ Ending value

$BV =$ Beginning value

$n =$ Number of years

Some State-of-the-art ICs

IC Name	Year	IC type & nm fabrication tech	Billion transistors	Million transistors /mm2	Processing speed per watt
A15 bionic by Apple for use in iPhone 13 Pro	2021	SoC, 5nm (SystemOnChip), 1 die	15	**138.9** =15,000/108	**0.18** =1.5TFLOPS/8.5W FP32/TDP (PL1)
M1 Ultra by Apple for desktop and notebook use	2022	SoC, 5nm (SystemOnChip), 4 dies	114	**131.9** =114,000/864	**0.35** =21.2TFLOPS/60W FP32/TDP (PL1)
H100 by Nvidia for Graphics and AI use	2022	GPU, 4nm (Graphics prosessing Unit), 1 die	80	**98.3** =80,000/814	**0.096** =67TFLOPS/700W FP32/TDP (?)
D1 (Dojo) by Tesla specifically for AI use	2022	NPU, 7nm (Neural processing Unit), 1 die	50	**77.5** =50,000/645	**0.057** =22.6/400W FP32/TDP (?)
Wafer Scale Engine 2 by Cerebras for AI	2020	NPU 7nm (Neural processing Unit) 1 die	2,600	**56.2** =2,600,000/46,225	**0.021** =503TFLOPS/23000W
V-NAND 3D chip by Micron used in MicroSD cards & SSDs (16Tbit=2TByte 2,666B transistors/TB)	2022	Flash memory 3D stacked chip, 232 layers per die, 16 dies stacked = 3712 layers in total, likely 16nm because 21/layer	5,333	**77,900** =5,333,000/68.5 **21.0** per layer =77,900/3,712	NA

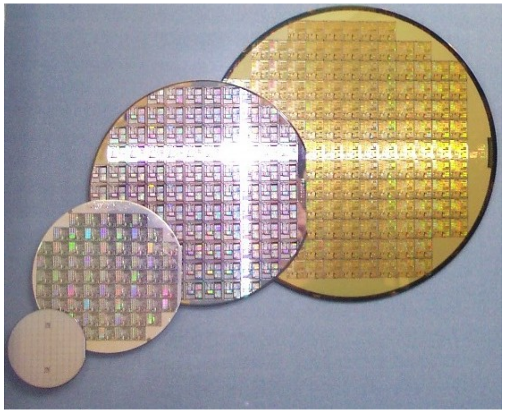
Sources for previous table

- **Apple A15:** https://en.wikipedia.org/wiki/Apple_silicon
- and https://www.cpu-monkey.com/en/cpu-apple_a15_bionic_5_gpu
- **Apple M1 Ultra (64 Core):** https://en.wikipedia.org/wiki/Apple_silicon#Apple_M1_Ultra
- **And** https://www.cpu-monkey.com/en/igpu-apple_m1_ultra_64_core-317
- And https://www.cpu-monkey.com/en/cpu-apple_m1_ultra_64_gpu
- **H100 by Nvidia:** https://www.nvidia.com/en-us/data-center/h100/
- and https://developer.nvidia.com/blog/nvidia-hopper-architecture-in-depth/
- and https://www.guru3d.com/news-story/nvidia-will-manufacture-h100-gpus-using-tsmc-4-nm-process.html
- **D1 (Dojo) by Tesla:** https://www.youtube.com/watch?v=ODSJsviD_SU&t=2566s (2:09:55)
- **Wafer Scale Engine 2:** https://f.hubspotusercontent30.net/hubfs/8968533/WSE-2%20Datasheet.pdf
- And https://www.marktechpost.com/2022/05/03/totalenergies-utilize-the-cerebras-cs-2-system-to-turn-an-ai-problem-long-accepted-to-be-memory-bound-into-compute-bound/
- **V-NAND chip by Micron:** https://en.wikipedia.org/wiki/Transistor_count
- and https://www.anandtech.com/show/17509/microns-232-layer-nand-now-shipping

Context: Ignots and Wafers

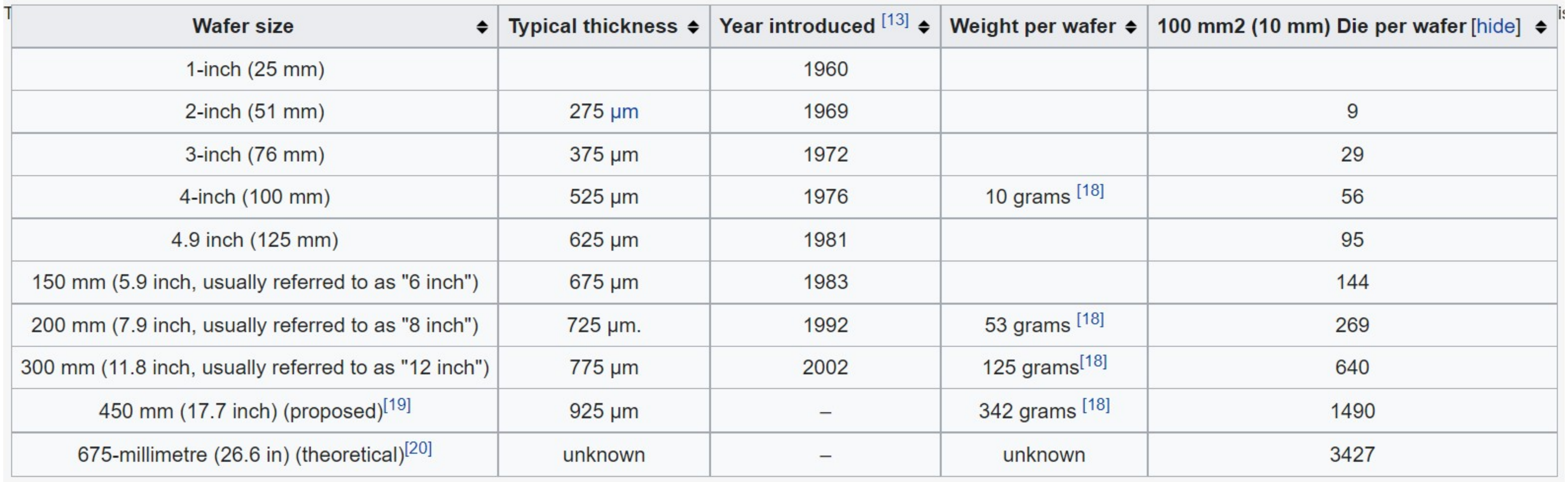
The silicon wafer as the material of semiconductor devices

A silicon wafer is a very thin round disk cut from high-purity silicon. A round silicon ingot is sliced to thicknesses of approximately 1mm. The surfaces of the resulting disk are polished carefully, then it is cleaned, resulting in the completed wafer. From this silicon wafer material, semiconductor devices are created.

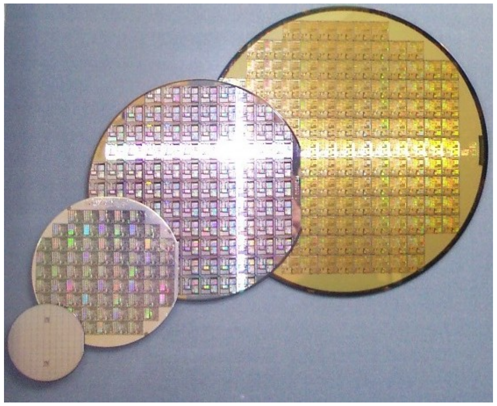


- **Sources:** https://www.sumcosi.com/english/products/about.html
- And https://en.wikipedia.org/wiki/Wafer_(electronics)

Context: Wafer size and die cuts

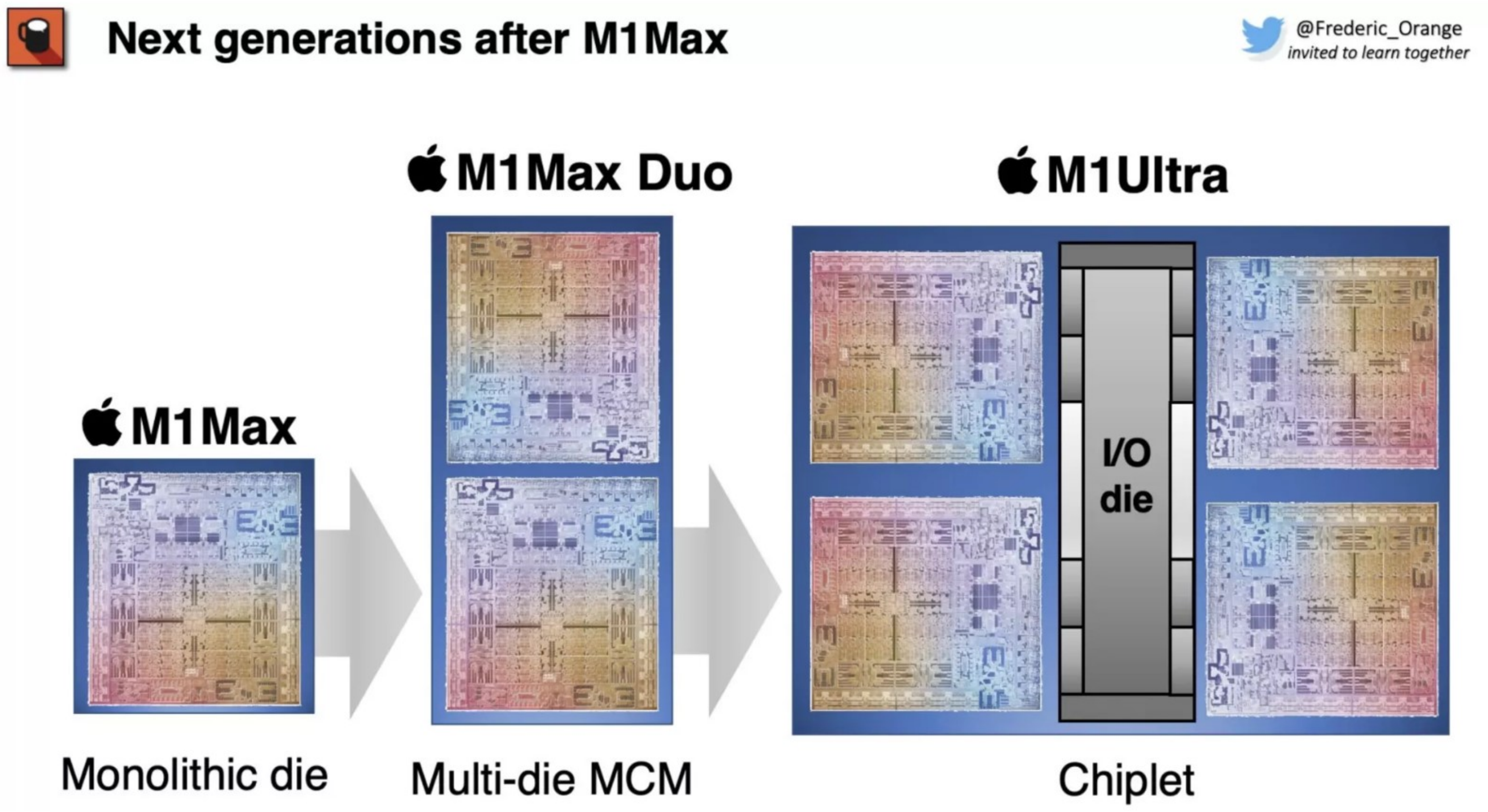


Wafer size	Typical thickness	Year introduced [13]	Weight per wafer	100 mm2 (10 mm) Die per wafer [hide]
1-inch (25 mm)		1960		
2-inch (51 mm)	275 μm	1969		9
3-inch (76 mm)	375 μm	1972		29
4-inch (100 mm)	525 μm	1976	10 grams [18]	56
4.9 inch (125 mm)	625 μm	1981		95
150 mm (5.9 inch, usually referred to as "6 inch")	675 μm	1983		144
200 mm (7.9 inch, usually referred to as "8 inch")	725 μm.	1992	53 grams [18]	269
300 mm (11.8 inch, usually referred to as "12 inch")	775 μm	2002	125 grams[18]	640
450 mm (17.7 inch) (proposed)[19]	925 μm	–	342 grams [18]	1490
675-millimetre (26.6 in) (theoretical)[20]	unknown	–	unknown	3427



- **Source:** https://en.wikipedia.org/wiki/Wafer_(electronics)

Context: Mono-die and Multi-die

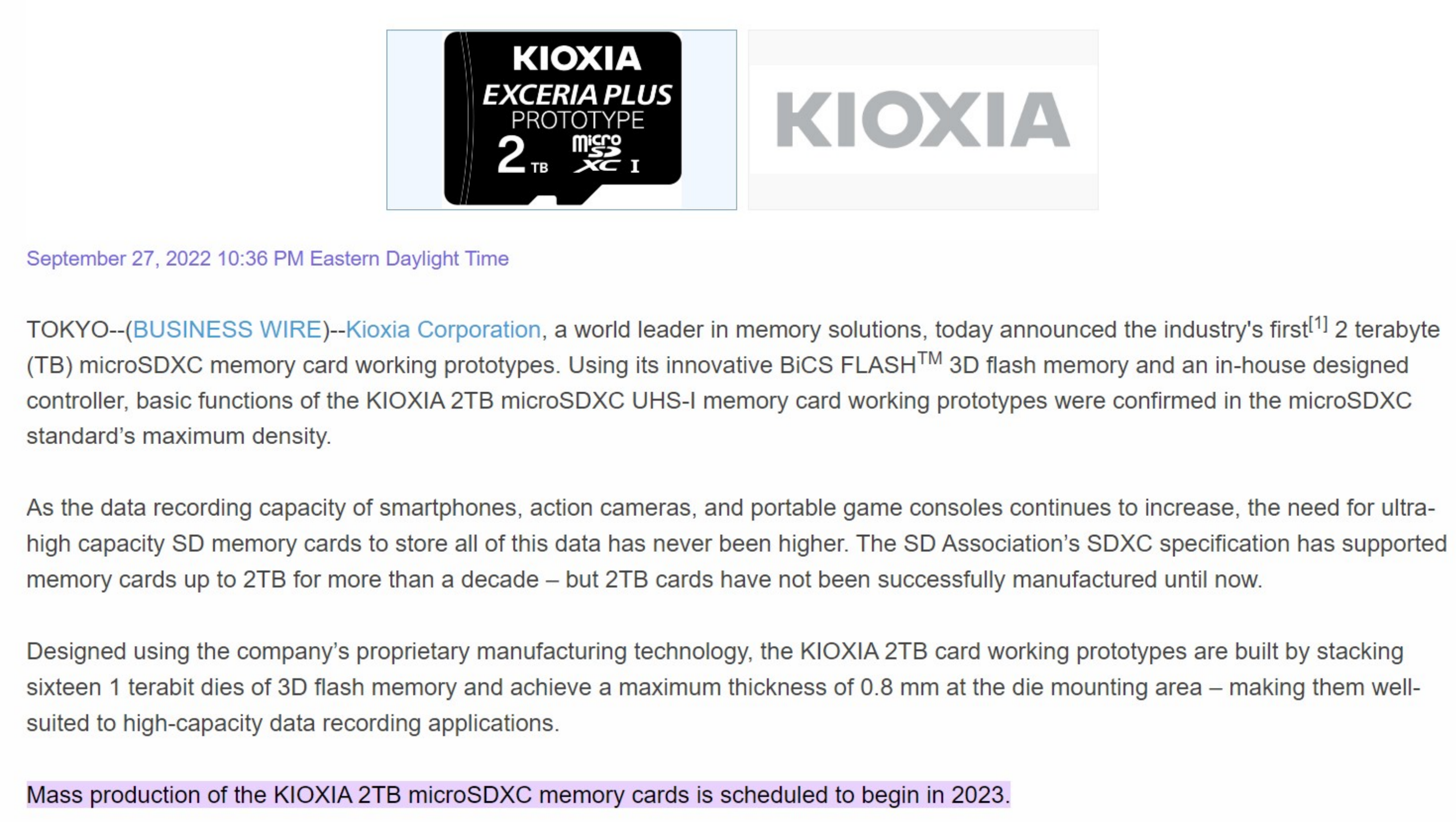


Next generations after M1Max

@Frederic_Orange
invited to learn together

- **Source:** https://twitter.com/frederic_orange?lang=en

The 2TB micro sd card: Mass market 2023



September 27, 2022 10:36 PM Eastern Daylight Time

TOKYO--(BUSINESS WIRE)--Kioxia Corporation, a world leader in memory solutions, today announced the industry's first[1] 2 terabyte (TB) microSDXC memory card working prototypes. Using its innovative BiCS FLASH^{TM} 3D flash memory and an in-house designed controller, basic functions of the KIOXIA 2TB microSDXC UHS-I memory card working prototypes were confirmed in the microSDXC standard's maximum density.

As the data recording capacity of smartphones, action cameras, and portable game consoles continues to increase, the need for ultra-high capacity SD memory cards to store all of this data has never been higher. The SD Association's SDXC specification has supported memory cards up to 2TB for more than a decade – but 2TB cards have not been successfully manufactured until now.

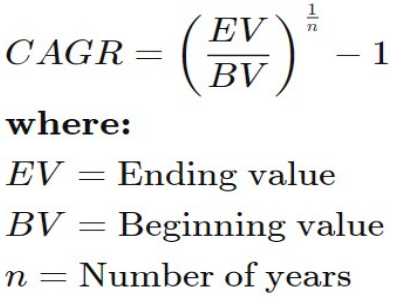
Designed using the company's proprietary manufacturing technology, the KIOXIA 2TB card working prototypes are built by stacking sixteen 1 terabit dies of 3D flash memory and achieve a maximum thickness of 0.8 mm at the die mounting area – making them well-suited to high-capacity data recording applications.

Mass production of the KIOXIA 2TB microSDXC memory cards is scheduled to begin in 2023.

- **Source:** https://www.businesswire.com/news/home/20220927006137/en/Kioxia-Develops-Industry%E2%80%99s-First-2TB-microSDXC-Memory-Card-Working-Prototypes#:~:text=Mass%20production%20of%20the%20KIOXIA,scheduled%20to%20begin%20in%202023.&text=%5B1%5D%20As%20of%20September%2028%2C%202022.

Data storage ICs not compute ICs are the king of transistor density

- ✓ **In 1965 and 1975** when Moore stated his “Law” transistor based ICs for long-term data storage did not exist
- ✓ **Today data storage ICs** are obviously the king of transistors per IC so we need to redo calculation including those to check Moore’s Law
- ✓ **State-of-the-art 2TB micro sd cards with 5.3 trillion transistors (using the Micron 16 dies chiplet) will go into mass production in 2023**
- ✓ **From 1971 to 2023 is 52 years**
- ✓ **The compounded annual growth rate (CAGR) can thus be calculated as**
- ✓ **((5,333,000,000,000/5000)^(1/52))-1= 0.491**
- ✓ **In percentage growth per year 49.1%**
- ✓ **Growth in 2 years is calculated as**
- ✓ **1*(1+0.491)^2=2.224. If 10 years you get 54.470**
- ✓ **2.224 is still about 2 so Moore’s law is spot on with a doubling every second year in transistors per chip. Note at 10 years differences are big**



$$CAGR = \left(\frac{EV}{BV}\right)^{\frac{1}{n}} - 1$$

where:

EV = Ending value

BV = Beginning value

n = Number of years

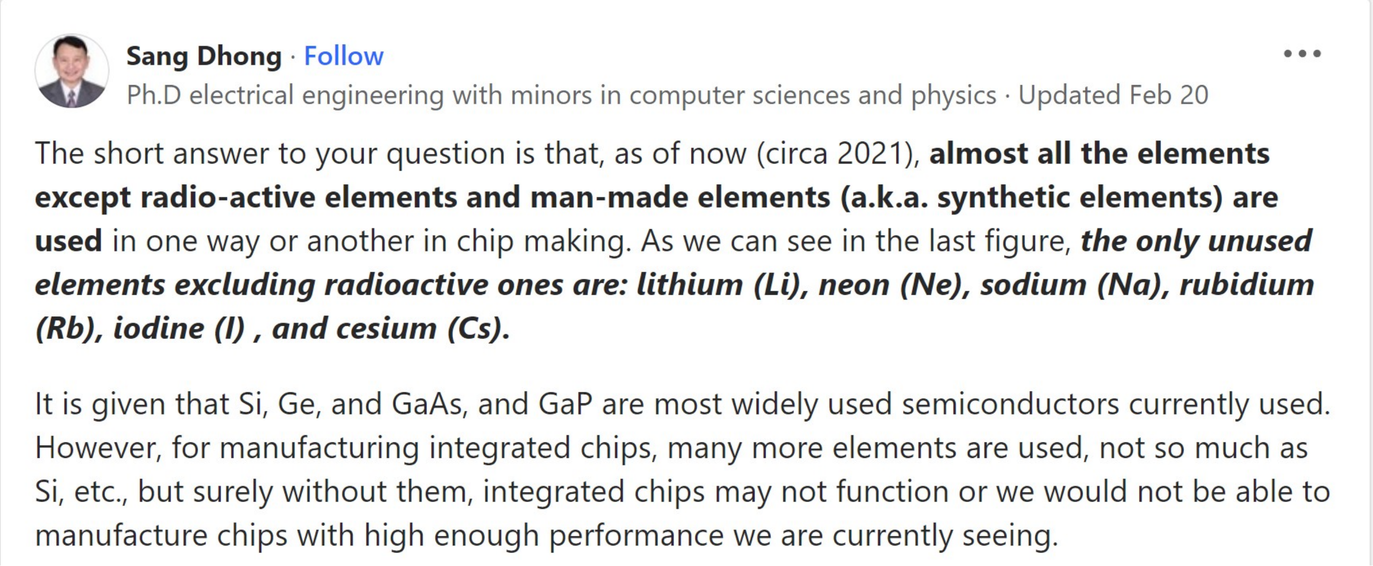
When Will Moore's Law End?

- ✓ **Moore's Law will end** when we can no longer shrink the size of transistors build on ICs
- ✓ Obviously transistors cannot be any smaller than the size of the atoms they are made of
- ✓ **So the questions become 1) how big are the atoms used to make transistors and 2) how many atoms do we need to make a functioning transistor currently and at the end of Moore's Law?**

Context: Size of atoms used for IC making by nm

Atom / element	Size of atom in nm	# of atoms to fit within 1 nm
Silicon	0.220 nm	5=1/0.22
Hydrogen	(smallest of all atoms) 0.050 nm	20=1/0.05
Cesium	(largest of all atoms) 0.520 nm	2=1/0.52

Asked at Quora.com: What are other elements used in chip making other than silicon?



Sang Dhong · Follow

Ph.D electrical engineering with minors in computer sciences and physics · Updated Feb 20

The short answer to your question is that, as of now (circa 2021), **almost all the elements except radio-active elements and man-made elements (a.k.a. synthetic elements) are used** in one way or another in chip making. As we can see in the last figure, ***the only unused elements excluding radioactive ones are: lithium (Li), neon (Ne), sodium (Na), rubidium (Rb), iodine (I) , and cesium (Cs).***

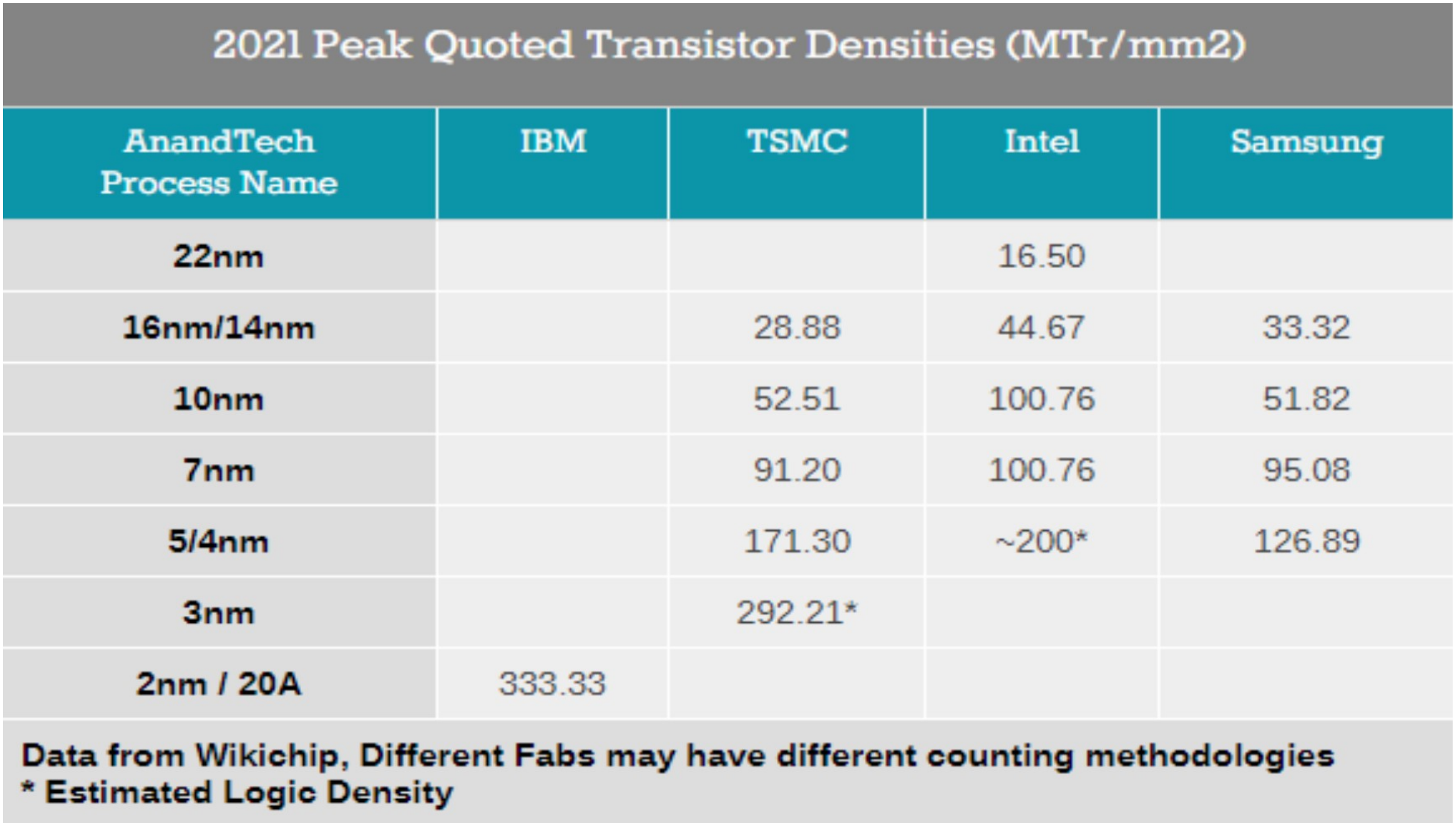
It is given that Si, Ge, and GaAs, and GaP are most widely used semiconductors currently used. However, for manufacturing integrated chips, many more elements are used, not so much as Si, etc., but surely without them, integrated chips may not function or we would not be able to manufacture chips with high enough performance we are currently seeing.

Si=Silicon
Ge=Germanium
GaAs=Gallium Arsenic
GaP=Gallium Phosphorus

Sources for previous table

- **Size of any atom:** https://en.wikipedia.org/wiki/Atomic_radii_of_the_elements_(data_page) (I use the reported empirial radius*2 and translated to nm)
- **Quora.com:** https://www.quora.com/What-are-other-elements-used-in-chip-making-other-than-silicon

Context: nm and transistor densities



2021 Peak Quoted Transistor Densities (MTr/mm2)

AnandTech Process Name	IBM	TSMC	Intel	Samsung
22nm			16.50	
16nm/14nm		28.88	44.67	33.32
10nm		52.51	100.76	51.82
7nm		91.20	100.76	95.08
5/4nm		171.30	~200*	126.89
3nm		292.21*		
2nm / 20A	333.33			

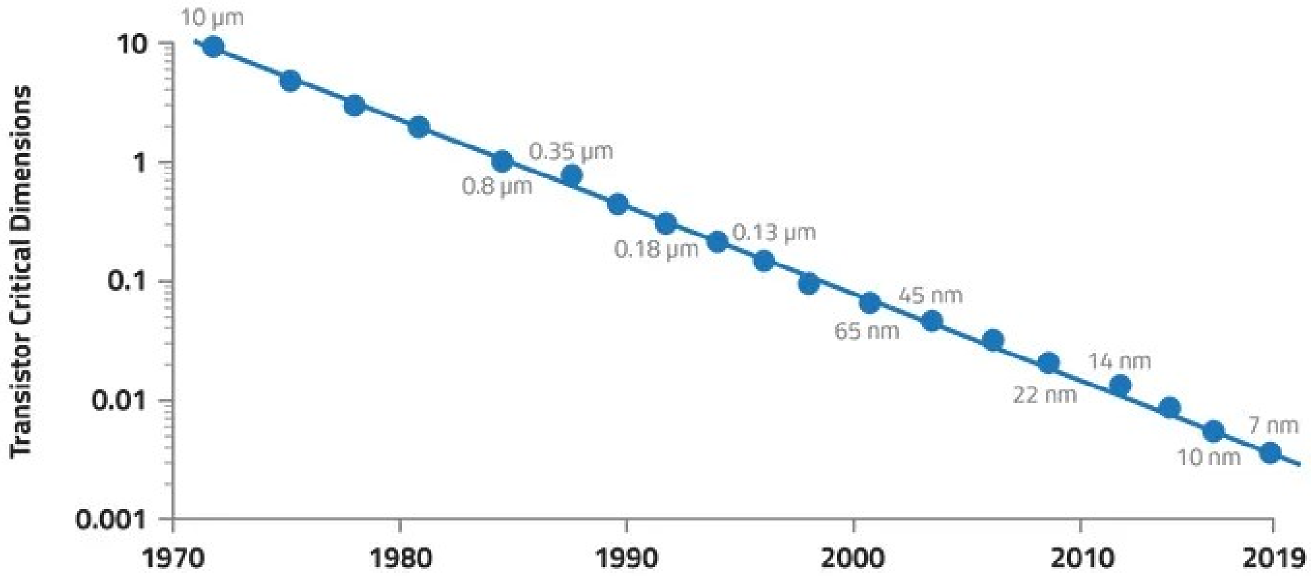
Data from Wikichip, Different Fabs may have different counting methodologies
*** Estimated Logic Density**

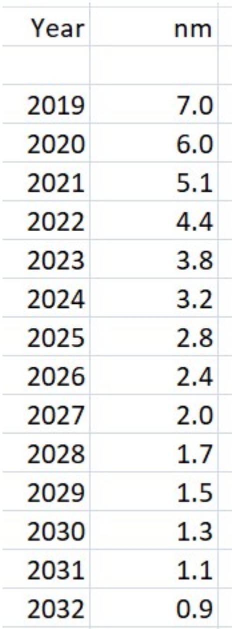
- **Source:** https://www.anandtech.com/show/16656/ibm-creates-first-2nm-chip

The Smallest Possible IC Structures

- From previous slide we can see atoms range from 0.05 to 0.5 nm
- In order to build the tiniest structure physically possible you will likely need at least 5 atoms side by side
- **Therefore, depending on which atom we consider we should be able to build circuit structures in the future that are between 0.25 and 2.5 nm large and on average about 1 nm but not any smaller than that**
- **To figure out when we will reach 1nm tech** we should study if there is any regularity by which new IC production tech is made available commercially

Progression of IC production tech: “nm Law”





Year	nm
2019	7.0
2020	6.0
2021	5.1
2022	4.4
2023	3.8
2024	3.2
2025	2.8
2026	2.4
2027	2.0
2028	1.7
2029	1.5
2030	1.3
2031	1.1
2032	0.9

- **Source:** https://semiengineering.com/scaling-up-and-down/ and https://en.wikipedia.org/wiki/Transistor_count
- **Graph shows** IC production tech decreases at a constant exponential rate because straight line fit on log scale
- Lets calculate it: We got 10 µm = 10,000 nm in 1972 and 7 nm in 2019
- **The compounded annual growth rate (CAGR)** can thus be calculated as
- **((7/10,000)^(1/47))-1= -0.143**
- So in percentage growth per year -14.3%
- **Using that we calculate table to the right. And we get 2031/2032 as the year Moore’s law ends regarding computational ICs**

1nm tech at 1 billion / mm2 by 2031

✓ **From slide 16 above we saw that IBM in 2021** announced they had build a proof-of-concept circuit using 2nm technology with a transistor density of 333 million per mm2

✓ To go from proof-of-concept to mass market production takes several years

✓ **The “nm Law ” derived above predicts that in 2027** we will see mass market applications such as smart phones build using 2nm ICs

✓ **That nm Law also predict that by 2031/2032** we will see mass market ICs build with 1 nm technology

✓ **It is reasonably to expect that the 1 nm mass market computational ICs made in 2031 will have a transistor density of about 1 billion transistors per mm2**

✓ **The logic of assuming 1 billion transistors per mm2 for 1nm tech** is that you can fit 4, 1nm squares on each 2nm square so multiply the 333 million transistors IBM got on their 2nm tech by 4 and then subtract some overhead and we get about 1 billion transistors

✓ **Note that 1 billion transistors / mm2 is about 10x of what we got in a new iPhone in 2022**

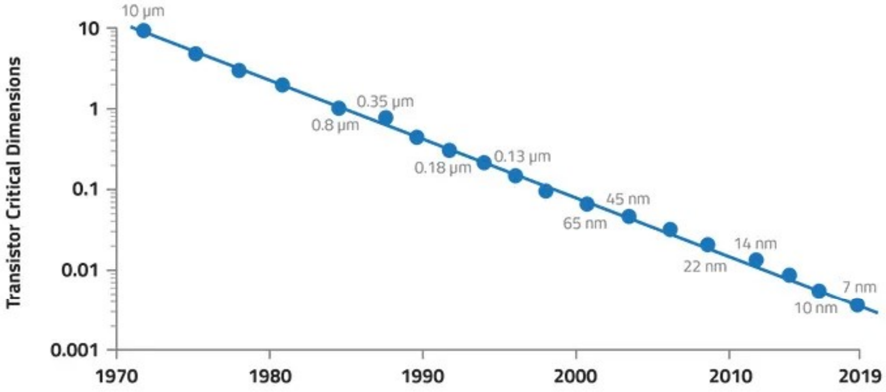
Data-storing ICs can be stacked

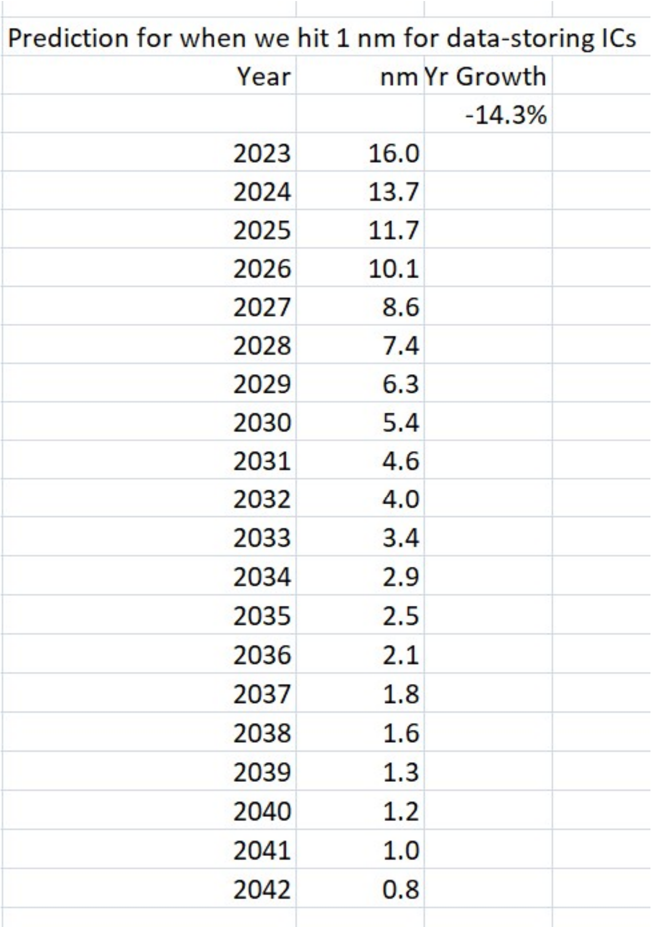
- ✓ However, 2031 is not the end for Moore’s Law regarding **data-storing ICs**
- ✓ **Unlike, computational and electronic ICs, data-storing ICs can be stacked** in multiple layers because their thermal/heat creation are many orders of magnitude less than that of computational ICs
- ✓ **Computational electronic ICs cannot be stacked because the extra heat** generated by adding another layer of circuits on top of the first would lead to self destruction and meltdown when turned on
- ✓ **Note that slide 6 above** showed that computational ICs use between 0.1 to 0.8 watt per mm2 to run or nearly 10 to 80 watt in 1 cm2 (a thumbnail) a few nms thick!!
- ✓ **This is a lot of energy in a very small space** so no wonder heat is a big issue for computational ICs

End of Moore's Law for Data-storing ICs?

- ✓ **To estimate when Moore's law will end for data-storing ICs** note that slide 6 and 11 showed mass market state-of-the art Micron based microSD cards will be available in 2023 using 16 nm tech and a single layer transistor density of 21 million/mm2
- ✓ **Assuming 5 years between each 50% reduction in nm production tech** it will take 20 years (16,8,4,2,1) from 2023 or until 2043 before data-storing ICs reach the 1nm minimum limit with a transistor density of 1 billion per mm2
- ✓ **We can also use the previously deducted "nm Law"** to calculate the time Moore's law will end for data-storing ICs
- ✓ **This "nm Law" will give us 2041** as the year Moore's Law end for data-storing ICs which is close enough to consider the above assumptions validated as likely to be true

Progression of IC production tech: “nm Law”





Prediction for when we hit 1 nm for data-storing ICs		
Year	nm	Yr Growth
		-14.3%
2023	16.0	
2024	13.7	
2025	11.7	
2026	10.1	
2027	8.6	
2028	7.4	
2029	6.3	
2030	5.4	
2031	4.6	
2032	4.0	
2033	3.4	
2034	2.9	
2035	2.5	
2036	2.1	
2037	1.8	
2038	1.6	
2039	1.3	
2040	1.2	
2041	1.0	
2042	0.8	

- **Slide 18 showed how to calculate** percentage growth per year -14.3%
- **Using that we calculate table above to the right.**
- **And we get 2041 as the year Moore’s law ends for data-storing ICs**

16 trillion transistors / mm3 by 2043

- ✓ **In order to estimate maximum transistor density per mm3 (cubic mm) in 2043** we would first need to take the square root of 1 billion transistors
- ✓ 31,622=(1,000,000,000)^0.5
- ✓ **That compares to 31,622 layers per mm**. Assuming an overhead of 50% for supporting structures (dies, cooling and other stuff) it could fit 15,811 layers of 1 billion transistors per layer into one mm3
- ✓ **Therefore, end of Moore's Law for data-storing ICs** would likely be 2043 with a transistor density of 15.811 trillion transistors per mm3 or 15,811 trillion transistors per cm3!
- ✓ Slide 6 above shows we need 2.666 trillion transistors to store 1TByte so **a cm3 of data-storing ICs in 2043 should be expected to store 5,930 TByte**= (15,811/2.666)

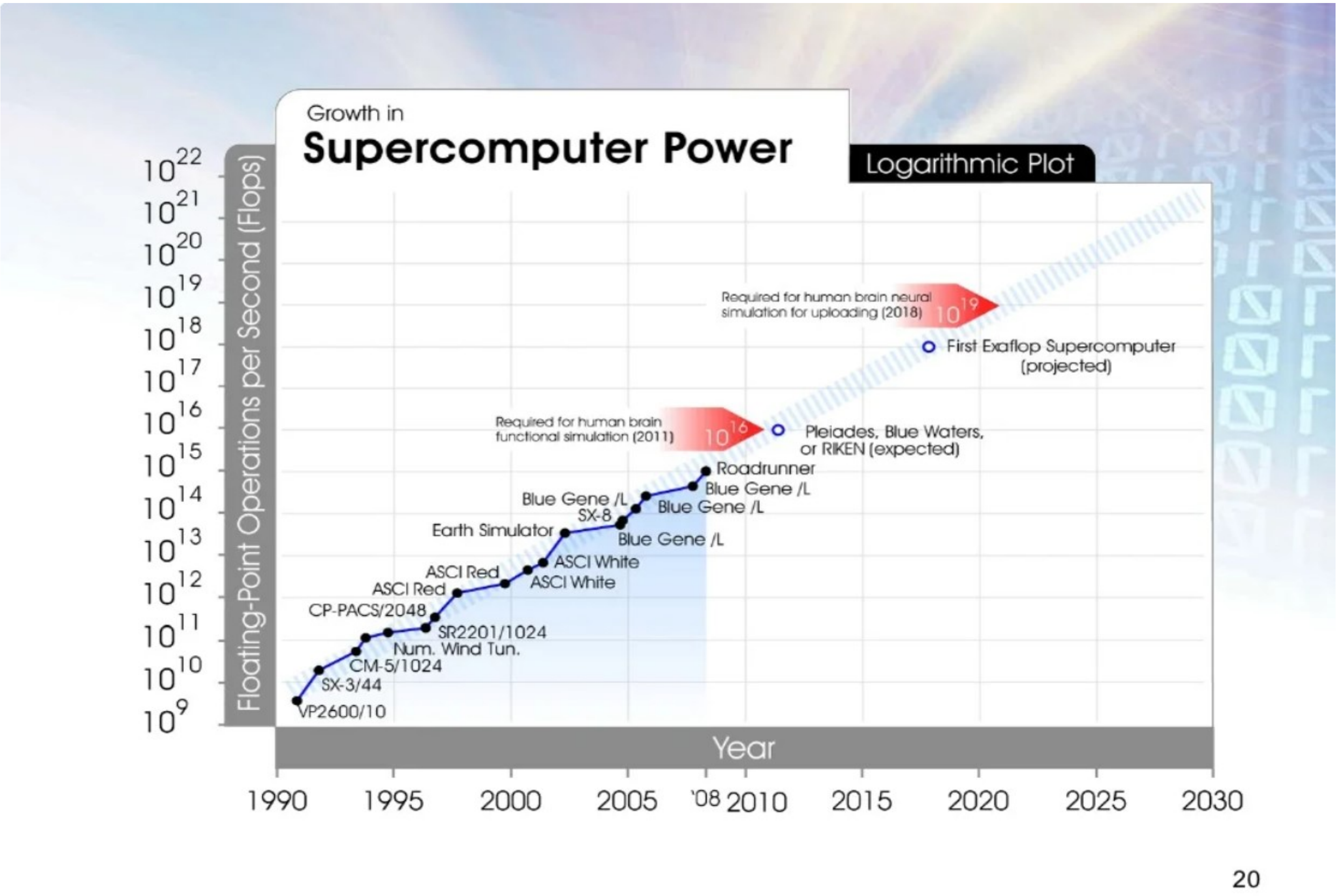
Comparisons: Data-storing ICs

- ✓ **For comparison, a typical iPhone 14 has 0.25TByte of date storage in 2022** so about 24,000 times less storage than should be possible for a new iPhone in 2043 using 1cm3 of 1nm data-storing ICs!
- ✓ **Moreover, the human brain has been estimated to store 2.5 petaByte of data** or 2,500Tbytes of data. This number is what Google give you right up when searching
- ✓ **It compares to 300 years of video recording** why I think this number to much too high as no human can remember their lives in detail second by second as a video can
- ✓ Even if we stay with 2,500Tbytes a human brain will still store less than half of the data that 1cm3 of data-storing ICs should be expected to do in 2043
- ✓ The human brain is on average 1,200 cm3 large so **data storing capacity of ICs in 2043 will exceeds that of a human by at least 2,846 times per cm3**= (5,930 Tbyte/(2,500 Tbyte /1200))
- ✓ **Moreover, ICs store information with 100% accuracy for decades** whereas brains are grossly erroneous at storing information and also constantly lose that information
- ✓ **To defend the human brain it also do computation in addition to data storage** within its 1200cm3 of space so not a fair comparison
- ✓ **Indeed, computation is truly where the human brain excels**

Comparisons: Computational ICs

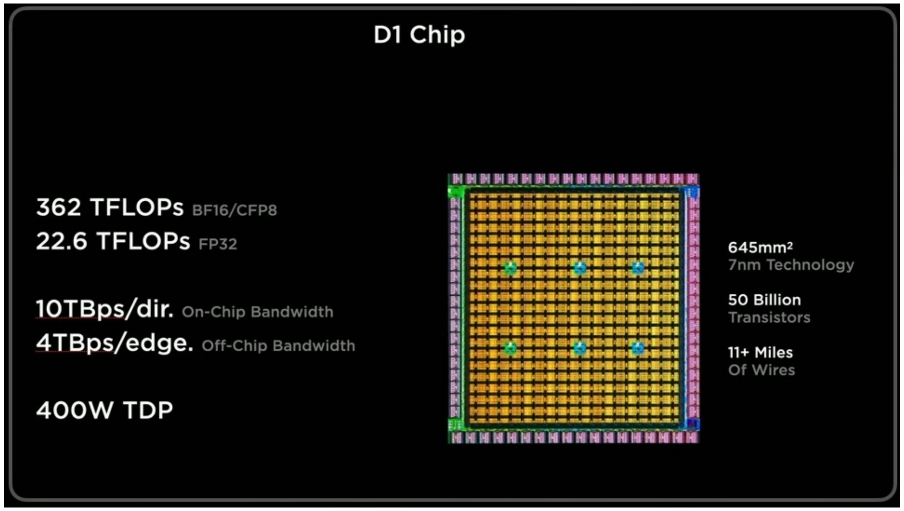
- ✓ Kurzweil [2009] estimated **the human brain can do about 10,000,000 TFLOPS (10 million trillion calculations per second=10exaFLOPS) using only 20W. That is 500,000 TFLOPS/Watt!**
- ✓ In slide 6 above we observed only **0.35 TFLOPS/Watt for best computational IC in 2022** (Apple M1 Ultra)
- ✓ **Moore's Law will end for computational ICs in about 2031** with a roughly 10X improvement in transistor density and therefore also ability to parallel compute in one IC
- ✓ That is only **3.5 TFLOPS/W** which is still far below the computational efficiency of the human brain
- ✓ **Another 16X can be gained from better computational designs** regarding the very simple computations that neural networks require and that compares much better to the kind of computation that happens in a human brain that is a neural network by design
- ✓ Specifically, you gain about 16X in FLOPS speed by calculating using BF16/CFP8 rather than the FP32 standard that was used in slide 6 above
- ✓ **A BF16 calculation has less precision than a FP32** but that is not a limiting factor for the AI training that nearly all new supercomputers are designed to do well
- ✓ So, expect computational ICs to max out at about **56** (=**3.5*16) TFLOPS/Watt in 2031, still much less than the perhaps 500,000 TFLOPS/W for human brain**

Kurzweil presentation [2009, slide 20]



- **Fastest supercomputer in 2022 is 1.1ExaFLOPS (likely BF16): Source:** https://en.wikipedia.org/wiki/List_of_fastest_computers
- **Source:** https://www.slideshare.net/antonioeram/raymond-kurzweil-presentation

Tesla AI day 2022



D1 Chip

362 TFLOPs BF16/CFP8
22.6 TFLOPs FP32

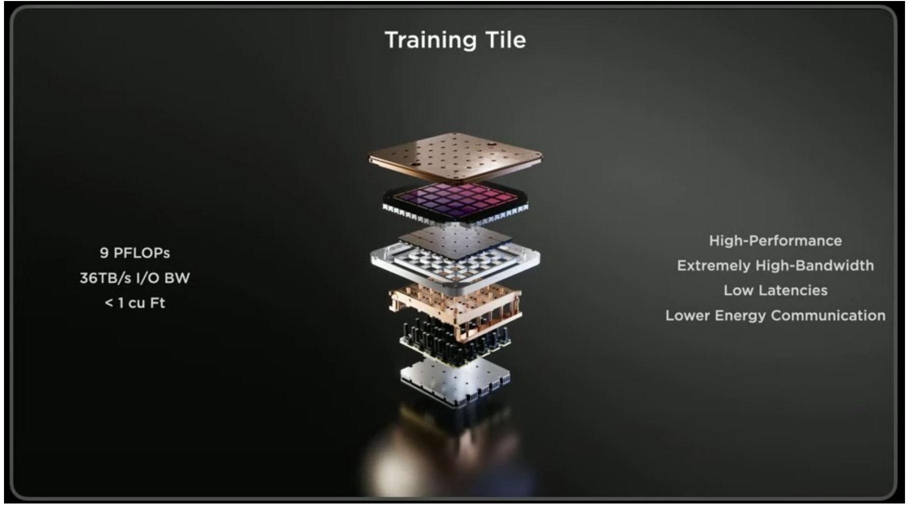
10TBps/dir. On-Chip Bandwidth
4TBps/edge. Off-Chip Bandwidth

400W TDP

645mm²
7nm Technology

50 Billion
Transistors

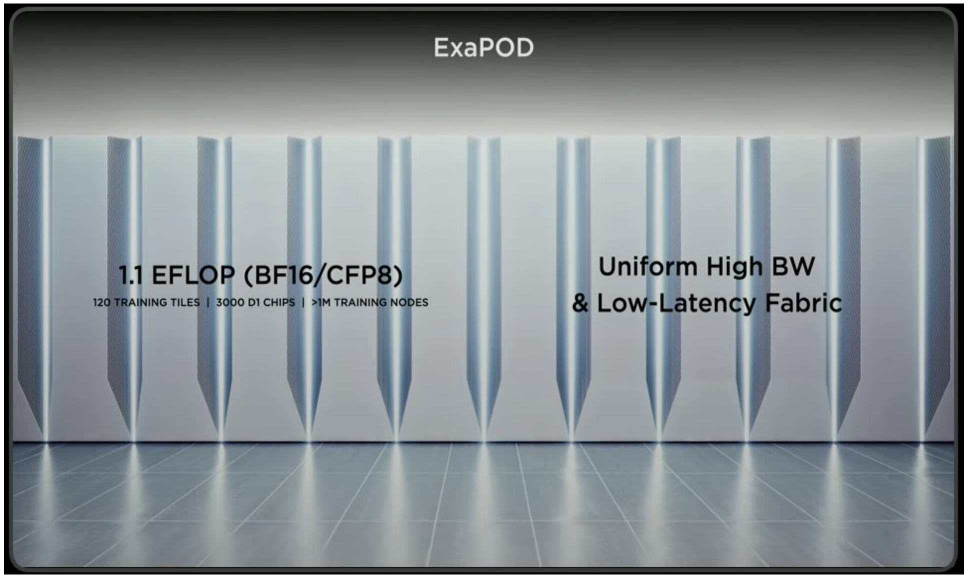
11+ Miles
Of Wires



Training Tile

9 PFLOPs
36TB/s I/O BW
< 1 cu Ft

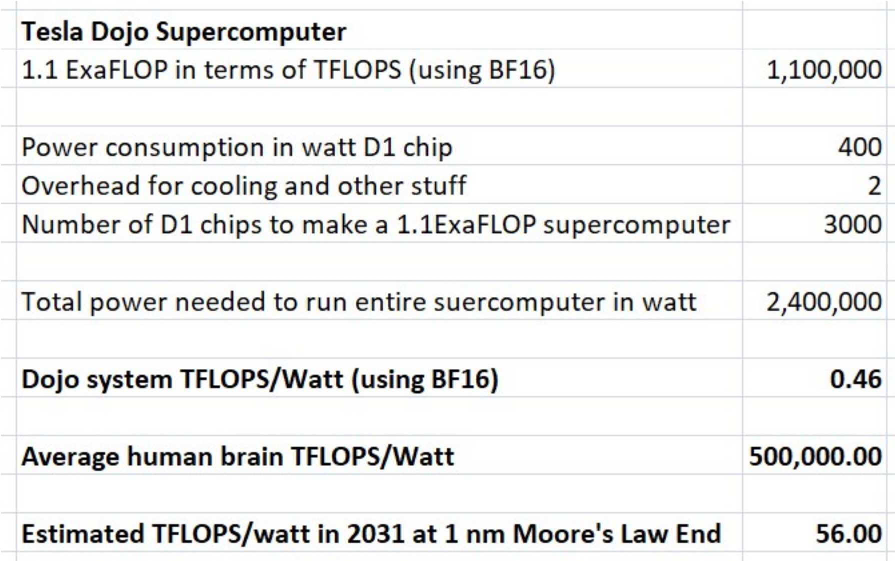
High-Performance
Extremely High-Bandwidth
Low Latencies
Lower Energy Communication



ExaPOD

1.1 EFLOP (BF16/CFP8)
120 TRAINING TILES | 3000 D1 CHIPS | >1M TRAINING NODES

Uniform High BW
& Low-Latency Fabric



Tesla Dojo Supercomputer	
1.1 ExaFLOP in terms of TFLOPS (using BF16)	1,100,000
Power consumption in watt D1 chip	400
Overhead for cooling and other stuff	2
Number of D1 chips to make a 1.1ExaFLOP supercomputer	3000
Total power needed to run entire suercomputer in watt	2,400,000
Dojo system TFLOPS/Watt (using BF16)	**0.46**
Average human brain TFLOPS/Watt	**500,000.00**
Estimated TFLOPS/watt in 2031 at 1 nm Moore's Law End	**56.00**

- **Source:** https://www.youtube.com/watch?v=ODSJsviD_SU&t=2566s (2:09:55)
- TDP is Thermal Design Power

Better efficiency with optical computational ICs

- ✓ **Fortunately, electronic ICs are not the only kind of ICs** that are able to compute data
- ✓ **The obvious alternative is to use optical ICs** that use photons instead of electrons to do the calculations
- ✓ **It is obvious because** photons 1) **need less energy than electrons to moved around** and 2) **they move faster than electrons** and 3) can **apply numerous of wavelengths simultaneously for each optical transistor** thereby doing the work of multiple electronic transistors in a single optical transistor
- ✓ **Optical ICs should have the potential to become many orders of magnitudes more efficient** than electronic ICs at computing
- ✓ **To repeat, in order to beat the human brain in computational efficiency per watt** we need a 10,000X improvement over what we can expect to have in electronic computational ICs when they max out in about 2031 (500,000 TFLOPS/W v. 56 TFLOPS/W)

Sources for previous 5 slides

- **Human brain memory:** https://www.cnsnevada.com/what-is-the-memory-capacity-of-a-human-brain/#:~:text=As%20a%20number%2C%20a%20%E2%80%9Cpetabyte,2.5%20million%20gigabytes%20digital%20memory.
- **Human brain memory 2:** https://www.scientificamerican.com/article/what-is-the-memory-capacity/
- **Human brain computational power:** https://neurotray.com/how-many-calculations-per-second-can-the-human-brain-do/
- **Human brain computational power2:** https://aiimpacts.org/brain-performance-in-flops/ (this source is better. It shows there are not much scientific consensus about how many TFLOPS a human brain has and the estimates varies ranging from 10^{12} to 10^{28} FLOPS so between 1 TFLOPS and 10,000,000,000,000,000 TFLOPS! I have chosen to use 10,000,000 TFLOPS or 10^{19} FLOPS which is in line with Kurzweil [2009] and Sandberg and Bostrom [2008] see below. This could be grossly wrong obviously.
- **Human brain computational power3:** Raymond Kurzweil presentation [2009, slide 20] https://www.slideshare.net/antonioeram/raymond-kurzweil-presentation R. Kurzweil estimate 10^{19} FLOPS are needed to simulate a complete human brain in a supercomputer. I have trust in Kurzweil as he is well known globally and has been spot on with many predictions he has made on the progression of technology although he has also been wrong on a few.
- **Human brain computational power4:** Sandberg and Bostrom [2008] http://www.fhi.ox.ac.uk/brain-emulation-roadmap-report.pdf

Final observations: Another 10X might be doable

- ✓ **Note that even at 1 billion transistors per mm2 made using 1nm tech** each transistor on average covers 31.6 nm = (1mm/(1,000,000,000^0.5))*1,000,000 (nm in a mm)
- ✓ Measured in #of 1nm2 each transistor cover we get 31.6^2 =**1000 *1 nm2**
- ✓ **1000 nm2 is much larger than the 1nm tech** used to make the transistor
- ✓ **One reason is that each transistor is build by many tiny features** each of which could be made by 1nm tech at the 1nm scale
- ✓ **Another reason is that ICs do not only contain transistors** although that is by far the most frequent component
- ✓ **ICs also contain resistors, capacitors, diodes, etc and all the wires to connect them**
- ✓ Therefore, the 1000 nm2 are needed to include all these components that each are build on systems of 1nm features
- ✓ **My point is, there is a potential for further shrinking** of the space needed to build an average transistor using 1nm tech simply by developing better component designs for the ICs both with regard to the computational and the data storing kind
- ✓ **On this account there is perhaps a potential for a further 10X improvement** of transistor density so only 10nm^2 =100 *1 nm2 are needed per transistor
- ✓ 10X density would enable **10 billion transistors/mm2** or **1,581 trillion transistors/mm3**
- ✓ **To achieve this 10X improvement** may also take additional years after 2031 and 2043

Conclusions: 1 of 2

- ✓ **This presentation has argued that Moore's Law will end in 2031** with regard to computational and electronic ICs using 1nm fabrication tech
- ✓ Such ICs will have about **1 billion transistors/mm2** with a compute efficiency of **56 TFLOPS/Watt** (using BF16)
- ✓ This is 10X better transistor density and a 10X (usingFP32) to 160X (using BF16) better compute efficiency as measured from best computational ICs anno 2022 (usingFP32)
- ✓ **This presentation also argued that Moore's Law will end in 2043** with regard to the data-storing ICs using the 1nm fabrication tech
- ✓ Such ICs will have a transistor density of about **15.811 trillion transistors per mm3** and 1 cm3 of data-storing ICs in 2043 should be expected to store about 5,930 Tbyte
- ✓ **It has further been argued that there is a potential for an additional 10X improvement** from denser IC component designs at the 1nm level so up to 10 billion transistors per mm2 might be doable
- ✓ **For data-storage** that would imply a 100X increase or 1,581 trillion trans./mm3!!

Conclusions: 2 of 2

- **HM do not know of a better future technology emerging for data-storing after the data-storing ICs** likely max out in about 2043
- **However, it is obvious from first principles thinking that computational electronic ICs** will be superseded by optical ICs that has the potential to be many orders of magnitude more compute efficient than computational electronic ICs
- **Another possibility to increase the compute efficiency of ICs is to use quantum computers** that exploit quantum superpositions of particles to create quantum bits that subsequently are used to generate ultrafast computation
- **While quantum computing appears to have great potential** it has not yet reached a technological maturity where any practical/useful calculation problems has been solved using a quantum computer
- **On the other hand, optical ICs are already in production** and used in supercomputers to solve real world problems mostly in AI
- Expect a forthcoming video at hmexperience.dk about optical ICs/supercomputers
- You may subscribe to this video and others at the YouTube channel HMexperience