Global Computing FLOPs Sold: A Historical Analysis and Future Projections (1984-2044)

1. Executive Summary

This report charts the extraordinary expansion of global computing power, measured in the aggregate theoretical Floating-Point Operations per Second (FLOPs) of hardware sold annually, from 1984 to 2024, with projections to 2044. The analysis documents a multi-trillion-fold increase in FLOPs capacity entering the market each year. This growth trajectory has been marked by distinct architectural eras: initial dominance by mainframes and minicomputers, followed by the personal computer (PC) revolution largely driven by the x86 architecture, the subsequent mobile transformation powered by ARM, and the current period of hyper-exponential growth fueled by Graphics Processing Units (GPUs) and specialized accelerators optimized for Artificial Intelligence (AI).

Total annual FLOPs sold grew from an estimated level below one PetaFLOPS (1015 FLOPs) in 1984 to the YottaFLOPS (1024 FLOPs) scale by 2024. Key architectural transitions are clearly visible in the data. Mainframes and minicomputers provided the bulk of floating-point capability in 1984. By 1994, the sheer volume of PCs, overwhelmingly based on the x86 architecture, made them the dominant source of newly sold FLOPs, despite lower per-unit performance compared to workstations. The year 2004 witnessed the rise of powerful discrete GPUs for gaming, which began to contribute significantly to total FLOPs, alongside the continued growth of x86 PCs and the nascent mobile market powered by ARM. By 2014, the massive unit volumes of ARM-based smartphones and tablets made them a major contributor to aggregate client-side FLOPs, while discrete GPUs solidified their lead as the primary source of floating-point performance entering the market. The current 2024 landscape is fundamentally reshaped by AI; data center GPUs and other AI accelerators now contribute the vast majority (>90%) of total FLOPs sold, often leveraging lower-precision formats (e.g., FP16, BF16, FP8) which yield significantly higher FLOPs counts than traditional FP32/FP64 measures used for CPUs and client GPUs.

Looking ahead, projections indicate continued strong growth through 2034, primarily driven by sustained investment in AI training and inference infrastructure. Architectural diversification is expected to continue, with ARM challenging x86 in servers and PCs, and RISC-V gaining traction as an open standard alternative. However, substantial headwinds exist. The immense power consumption of large-scale AI compute poses a significant challenge, alongside the escalating costs of semiconductor manufacturing and geopolitical uncertainties surrounding supply chains. These factors make the 2044 outlook highly speculative, suggesting that while growth will likely continue, its rate may moderate unless significant breakthroughs in energy efficiency or new computing paradigms emerge.

2. Introduction

This report analyzes the historical trajectory and future prospects of global computing power deployment, quantified through the metric of Floating-Point Operations per Second (FLOPs). FLOPs measure the number of calculations involving decimal numbers (floating-point arithmetic) that a processor can perform each second.[1] This metric is particularly relevant for performance assessment in computationally intensive fields such as scientific simulation, engineering, financial modeling, graphics rendering, and, increasingly, artificial intelligence.[1]
The core metric used in this analysis is the estimated *aggregate theoretical peak FLOPs capacity* of complete computing hardware systems (e.g., mainframes, servers, PCs, mobile devices) and key accelerator components (primarily GPUs) *sold* globally within specific calendar years. This "FLOPs Sold" metric reflects the rate at which new computational capability enters the market. It is distinct from the *utilized FLOPs* (which depends on software and workload) and the total *installed base* of computing capacity (which represents cumulative sales minus retirements, often estimated over a multi-year hardware lifespan [4]).
For consistency across diverse architectures and historical periods, this report primarily references theoretical peak single-precision (FP32) or double-precision (FP64) FLOPs, depending on the architecture's typical use case and data availability. However, the recent explosion in AI-driven compute necessitates acknowledging the performance of specialized accelerators (GPUs, TPUs, NPUs) in lower-precision formats like FP16, BF16, and FP8.[5] These formats offer significantly higher throughput (FLOPs) for AI workloads compared to FP32/FP64, and their contribution is explicitly noted in the analysis for 2024 and beyond, as they represent the dominant share of FLOPs sold today.
Tracking the annual volume of FLOPs sold is significant for several reasons. It serves as a proxy for the rate of technological advancement in computing hardware.[9] It underpins the feasibility of increasingly complex applications, from weather modeling to drug discovery to large language models. Furthermore, the sheer scale of deployed FLOPs has profound implications for global energy consumption, data center infrastructure development, and the economics of the semiconductor industry.
This report examines the FLOPs sold landscape over six distinct points in time: 1984, 1994, 2004, 2014, and 2024, followed by projections for 2034 and 2044. For each period, the analysis breaks down the total FLOPs contribution by dominant computing architectures, tracing the evolution from the mainframe and minicomputer era through the rise of the PC (Intel x86 and competitors), the mobile revolution (ARM), and the current ascendancy of parallel processors (GPUs) and AI accelerators. The report begins by detailing the methodology and inherent challenges in such long-range estimation, followed by a chronological analysis of each decade, and concludes with future projections and strategic insights.

3. Methodology and Challenges

Estimating the total FLOPs sold globally across decades requires a multi-step approach fraught with challenges, necessitating approximations and careful consideration of data limitations.

Estimation Approach:

The core methodology involves:

1. **Identifying Dominant Compute Categories:** For each target year, identifying the primary categories of computing devices contributing significantly to the market (e.g., 1984: Mainframes, Minicomputers, Personal Computers; 2024: Servers, PCs, Smartphones, Tablets, Data Center Accelerators).
2. **Identifying Key Architectures/Models:** Within each category, identifying the most prevalent processor architectures or specific high-impact models (e.g., IBM 308X mainframes, DEC VAX minicomputers, Intel x86 CPU generations, ARM Cortex series, Nvidia/AMD GPU architectures).
3. **Estimating FLOPs-per-Unit:** Determining a representative theoretical peak FLOPs value for each key architecture/model in a given year. This relies on a hierarchy of sources:
   - Direct manufacturer specifications or benchmark data (e.g., Linpack [11], SPEC [35], Whetstone [40]).
   - Theoretical calculations based on architecture details (cores, clock speed, operations per cycle).[48]
   - Academic studies providing historical performance estimates (e.g., Nordhaus MSOPS [9], McCallum MIPS/SPEC data [35]).
   - Conversions from other metrics like MIPS (Millions of Instructions Per Second) when direct FLOPs data is unavailable, particularly for older systems (discussed below).
4. **Gathering Unit Sales Data:** Compiling annual global unit shipment data for each major category and, where possible, by architecture or key vendor. This relies heavily on market research firms (e.g., IDC, Gartner, Jon Peddie Research, Mercury Research) and historical industry reports.[52]
5. **Calculating Total FLOPs:** For each category/architecture, multiply the estimated units sold by the estimated average FLOPs-per-unit. Sum these values across all segments to arrive at the total estimated FLOPs sold for the year.

Challenges and Approximations:

Several significant challenges necessitate approximations throughout this analysis:

- **Data Scarcity and Consistency:** Comprehensive, globally consistent data for both unit sales and FLOPs performance is often unavailable, particularly for earlier decades, niche architectures, or specific component types like early GPUs.[9] Market research data from different firms can also vary due to differing methodologies.[105] This requires careful source selection, cross-referencing, estimation based on related data points (e.g., revenue, regional data), and interpolation or extrapolation where direct data is missing. Historical academic studies [9] provide valuable benchmarks but may use metrics (like MSOPS) that require further interpretation. Finding specific historical market reviews from publications like Datamation or Byte for the target years proved difficult.[57]

- **MIPS to FLOPs Conversion:** A major challenge, especially for the 1984 analysis, is converting MIPS ratings (commonly reported for older mainframes and minicomputers) to FLOPs.[2] There is no universal conversion factor, as the ratio depends heavily on the specific architecture, instruction mix of the benchmark used to derive MIPS, and the efficiency of floating-point execution units. Based on available data for VAX systems [45], a ratio of roughly 3.5-4 MWIPS (Millions of Whetstone Instructions Per Second, often used for minis) per MFLOPS (Millions of Floating-Point Operations per Second, double-precision) appears reasonable for that architecture. For IBM mainframes of the era, the relationship is less clear; a conservative ratio is assumed, acknowledging this significant uncertainty. Nordhaus suggests MSOPS is roughly equivalent to MIPS [50], but the MSOPS-to-FLOPS relationship is also not precise.[9]
- **Theoretical vs. Measured FLOPs:** This report uses theoretical peak FLOPs ratings where possible for consistency. It is crucial to understand that actual achieved performance in real-world applications is typically a small fraction (often less than 10%) of this peak value.[48] Factors like memory bandwidth limitations, cache misses, instruction dependencies, thermal throttling, and the specific instruction mix dramatically reduce sustained performance.[2]
- **Floating-Point Precision:** The definition of a "FLOP" depends on the precision of the numbers being calculated (e.g., 64-bit double-precision/FP64, 32-bit single-precision/FP32, 16-bit half-precision/FP16 or BFloat16, 8-bit FP8). While CPUs traditionally focused on FP64 and FP32, modern AI accelerators achieve significantly higher FLOPs counts by operating on lower-precision formats.[5] This report attempts to specify the precision basis (primarily FP32/SP or FP64/DP for historical systems and CPUs/client GPUs, shifting to mixed/lower precision for AI accelerators in 2024 and beyond) but comparisons across eras and architectures must account for these differences.
- **Defining "Sold":** The data often reflects system shipments (PCs, servers). However, especially for GPUs, component sales are a significant factor. This analysis primarily uses system-level sales data where available but incorporates discrete GPU component shipment data in later years as a necessary proxy for the FLOPs contribution from that segment. This introduces some ambiguity, as not all discrete GPUs sold end up in new systems within the same year.
- **Architectural Complexity:** Grouping diverse hardware into broad categories like "x86 PC" or "ARM Mobile" involves simplification. Performance within these categories varies widely based on specific models, clock speeds, and configurations (e.g., presence of a math coprocessor in early PCs [171]). Weighted averages based on estimated market mix are used, introducing further approximation.

Despite these challenges, this methodology provides a framework for estimating the scale and architectural trends of global computing power deployment over the past four decades and into the future. All figures presented should be understood as estimates subject to the limitations described.

4. 1984: The Mainframe/Minicomputer Era with PC Seeds

Market Context:
In 1984, the global computing landscape was characterized by the dominance of large-scale systems in enterprise and research environments. Mainframes, primarily from IBM, and minicomputers from vendors like Digital Equipment Corporation (DEC), Data General, and HP, represented the bulk of installed computing power and market value.52 However, the personal computer revolution was gaining momentum. While 8-bit home computers like the Commodore 64 and Apple II series enjoyed widespread popularity 53, the 16-bit IBM PC, XT, and the newly introduced PC/AT, based on Intel's x86 architecture, were rapidly establishing a standard in the business sector.175 Apple also made a significant impact with the launch of the Macintosh, featuring a graphical user interface and Motorola's 68000 processor.176 Notably, 1984 marked the first year that the revenue from desktop computer sales surpassed that of mainframes, signaling a major market shift.67
Unit Sales Estimates:

- **Mainframes:** Approximately 10,700 units were sold globally, with a market value exceeding $10 billion.[52] IBM's 308X series was a key product line.[182]
- **Minicomputers:** Global sales reached roughly 47,820 units, with a market value comparable to mainframes, around $10.5 billion.[52] DEC's VAX line, particularly the VAX-11/780, was highly influential.[42]
- **Personal Computers (PCs):** This category is more complex due to the distinction between business PCs and home computers. Data suggests ~2.2 million "micros" (primarily business-oriented PCs) were sold with a value of ~$7.9 billion.[52] However, total personal computer sales including home computers were significantly higher. The Apple II line sold approximately 1 million units [53], the Commodore 64 sold between 2 and 2.5 million units [185], and the new Macintosh sold around 250,000 to 370,000 units by year-end.[64] IBM PC, XT, AT, and PCjr sales, along with compatibles, likely accounted for another 2-3 million units.[175] Summing these major platforms suggests a total 1984 personal computer market size of approximately 6 to 7 million units. This analysis uses an estimate of 6.5 million total PC units sold.

FLOPs per Unit Estimates:
Estimating FLOPs for 1984 hardware requires careful interpretation of available metrics, often MIPS or benchmark results, and acknowledging the optional nature of floating-point hardware in PCs.

- **Mainframes (IBM 3081K):** Rated at ~7 MIPS.[183] Direct FLOPs ratings are scarce. The architecture supported double-precision floating-point.[182] Converting MIPS to MFLOPS for this architecture is highly uncertain.[167] Assuming a ratio where floating-point operations constitute a fraction of total instructions, a rough estimate might be 3-5 MFLOPS (DP).

- **Minicomputers (DEC VAX-11/780 with FPA):** Widely considered a 1 MIPS machine [42], it achieved ~1.02 MWIPS on the Whetstone benchmark.[42] More direct estimates place its double-precision floating-point performance at approximately 0.25 MFLOPS.[45] This value is used as representative for high-end minicomputers with FPAs.
- **PCs (Intel 8088/8086/80286):** The base CPUs (8088/86) lacked hardware floating-point capabilities. Performance relied on software emulation, likely delivering only 1-2 KFLOPS.[193] The optional Intel 8087 coprocessor provided ~0.05 MFLOPS (SP).[171] The IBM PC/AT with an 80286 and the newer 80287 coprocessor achieved roughly 0.009 MFLOPS (DP) according to one source [198], though estimates based on relative performance suggest potentially higher single-precision capability, perhaps ~0.1 MFLOPS (SP).[199] Crucially, the adoption rate of these expensive coprocessors was low, likely in the 10-15% range for systems sold in 1984, primarily in business or technical configurations.[56] Therefore, the *average* FLOPs per Intel-architecture PC sold was very low, estimated here at a weighted average of 0.008 MFLOPS (SP).
- **PCs (Motorola 68000 - e.g., Macintosh):** The 68000 CPU also lacked native floating-point instructions. Software emulation was similarly slow (~1-2 KFLOPS).[204] The Motorola 68881 FPU was announced in 1984 [204], offering ~0.16 MFLOPS (SP) at 16MHz, but its integration into systems sold within 1984 was likely minimal. The average FLOPs per Motorola-based PC is estimated at 0.001 MFLOPS (SP).
- **PCs (MOS 6502 - e.g., Apple II, C64):** These 8-bit processors had negligible floating-point capability, relying entirely on very slow software emulation.[193] Average FLOPs per unit is estimated at <<0.001 MFLOPS (SP), effectively zero for aggregate calculations.

Total FLOPs Calculation (Approximate):

- Mainframes: 10,700 units * 4 MFLOPS/unit ≈ 4.3 x 10^10 FLOPs (0.043 TFLOPS)
- Minicomputers: 47,820 units * 0.25 MFLOPS/unit ≈ 1.2 x 10^10 FLOPs (0.012 TFLOPS)
- PCs (Intel Arch): ~2.5M units * 0.008 MFLOPS/unit ≈ 2.0 x 10^10 FLOPs (0.020 TFLOPS)
- PCs (Motorola Arch): ~0.3M units * 0.001 MFLOPS/unit ≈ 3.0 x 10^8 FLOPs (0.0003 TFLOPS)
- PCs (MOS Arch): ~3.7M units * 0.0001 MFLOPS/unit ≈ 3.7 x 10^8 FLOPs (0.0004 TFLOPS)
- **Total Estimated FLOPs Sold in 1984:** ≈ 0.076 TFLOPS (or 76 GFLOPS)

Architectural Breakdown:

Based on the calculations above, mainframes contributed the largest share of total FLOPs sold (~57%), followed by Intel-architecture PCs (~26%), and then minicomputers (~16%). The contributions from Motorola and MOS-based PCs were negligible in terms of floating-point capability in 1984.

Table 1: Estimated 1984 FLOPs Sold by Architecture

Architecture Category	Est. Units Sold	Est. Avg. FLOPs/Unit (MFLOPS, SP/DP Mix)	Est. Total FLOPs (GFLOPS)	% of Total FLOPs

Mainframes (e.g., IBM 308X)	10,700	~4.0 (DP)	~42.8	~56.6%
Minicomputers (e.g., VAX)	47,820	~0.25 (DP)	~12.0	~15.8%
PC - Intel Arch (808x/286)	~2,500,000	~0.008 (SP, weighted avg w/ FPU)	~20.0	~26.4%
PC - Motorola Arch (68000)	~300,000	~0.001 (SP, software)	~0.3	~0.4%
PC - MOS Arch (6502)	~3,700,000	~0.0001 (SP, software)	~0.4	~0.5%
PC - Other	Minimal	Negligible	Negligible	~0.3%
Total	**~6,560,000**	**N/A**	**~75.5**	**100.0%**

(Note: All figures are estimates based on available data and stated assumptions, particularly regarding MIPS-to-FLOPS conversion and FPU attach rates. DP = Double Precision, SP = Single Precision.)

Analysis and Implications:

The 1984 data reveals a computing landscape in transition. While mainframes still delivered the highest per-system floating-point performance and contributed the majority of total FLOPs sold, the burgeoning PC market was already significant in terms of aggregate computational capacity entering the market. Despite the vastly lower FLOPs capability per individual PC compared to a mainframe or minicomputer (MFLOPS vs KFLOPS or less), the sheer volume of PC sales meant their collective contribution was substantial, already surpassing the total FLOPs from the minicomputer segment. This highlights the disruptive potential of volume economics; mass-market computing, even at lower individual performance levels, could rapidly increase the total available computational power.

However, the utility of these PC FLOPs for demanding scientific and engineering tasks was severely limited. The reliance on expensive, optional floating-point coprocessors like the Intel 8087 [171] or slow software emulation [193] created a significant bottleneck. This practical limitation meant that minicomputers and mainframes, with their integrated and more powerful floating-point capabilities, remained the platforms of choice for computationally intensive applications, despite the growing volume of PCs being sold.

Furthermore, the PC market itself was architecturally diverse. While IBM's adoption of Intel's x86 architecture was proving strategically significant [208], platforms based on Motorola's 68000 (Macintosh) and MOS Technology's 6502 (Apple II, Commodore 64) commanded enormous unit sales.[53] The slightly higher potential FLOPs contribution from the Intel segment, due to the availability of the 8087/287 FPU, offered a glimpse into the future x86 dominance in performance-sensitive PC applications, even as other architectures held larger shares of the overall personal computer market in 1984.

5. 1994: The x86 PC Era and Early GPU Seeds

Market Context:
A decade later, the computing landscape had transformed dramatically. The PC, predominantly based on the Intel x86 architecture and its compatibles (like AMD), running Microsoft Windows, had become the dominant computing platform globally.209 Mainframes continued to serve enterprise needs but represented a shrinking fraction of new compute capacity sold annually.68 The minicomputer market had largely given way to powerful RISC (Reduced Instruction Set Computing)-based workstations from vendors like Sun Microsystems (SPARC), DEC (Alpha), Silicon Graphics (MIPS), and HP (PA-RISC), which catered to high-performance technical and scientific computing. Apple had transitioned its Macintosh line from Motorola 68k to the PowerPC architecture, developed in partnership with IBM and Motorola.84 Concurrently, graphics accelerators (early GPUs) were becoming standard components in PCs, primarily focused on improving 2D Windows performance and enabling early 3D gaming, though their contribution to general-purpose FLOPs was still minimal compared to CPUs.214

Unit Sales Estimates:

- **Personal Computers (PCs):** Global shipments reached approximately 47.9 million units.[54] The market was led by Compaq (10.0%), Apple (8.3%), and IBM (8.2%).[54] The vast majority used x86 processors. Apple's sales included the newly launched Power Macintosh models; estimates suggest around 1 million PowerPC Macs may have shipped by the end of 1994.[84]
- **Servers and Workstations:** Precise combined unit data is scarce. The Unix workstation market was substantial, with Sun Microsystems being a major player.[71] Mainframe unit shipments were likely below 10,000 units.[62] The overall server market revenue was estimated at $141 million by IDC for a specific segment, which seems low for the total market [87]; total units were likely in the low millions (e.g., 1-2 million).
- **Graphics Processing Units (GPUs):** Virtually all PCs shipped included some form of graphics capability, either integrated onto the motherboard or via an add-in card. Key vendors for graphics chips included S3, ATI, Cirrus Logic, and Tseng Labs.[215] Nvidia existed but had not yet released its breakthrough products.[220] Estimating the attach rate of *discrete* GPUs (separate add-in cards, typically more powerful than integrated solutions) is difficult for this period, but it was likely lower than today, perhaps in the 30-40% range, suggesting around 15-20 million discrete units sold.

FLOPs per Unit Estimates:
Performance benchmarks like Linpack and SPEC became more common in this era, providing better (though still imperfect) FLOPs estimates.

- **Servers/Mainframes (e.g., IBM ES/9000):** A 6-processor ES/9000-720 system achieved 0.54 GFLOPS (DP) on Linpack.[14] A single processor's performance would be roughly 0.1 GFLOPS (DP).
- **Workstations (e.g., Sun SPARCstation 10/41, DEC Alpha 3000):** A SPARCstation

10/41 (SuperSPARC @ 40MHz) achieved ~8 MFLOPS (DP) on C Linpack.[16] A DEC Alpha 3000/700 (225MHz Alpha 21064A) achieved a SPECfp95 rating of 230.6.[35] Converting SPECfp ratios to MFLOPS is complex and benchmark-dependent [36], but suggests performance potentially in the 50-100 MFLOPS (DP) range for high-end Alphas. A reasonable average for RISC workstations sold in 1994 might be 20-40 MFLOPS (DP).
- **PCs (Intel 486DX2-66):** Linpack results indicate performance around 2.6 MFLOPS (DP).[12]
- **PCs (Intel Pentium 60/66 MHz):** Early Pentiums offered a significant floating-point improvement over the 486. SPECfp92 results were around 55-64.[35] Linpack results vary, but native code benchmarks suggest performance in the range of 10-20 MFLOPS (DP).[20] We estimate an average of 15 MFLOPS (DP) for Pentiums sold in 1994.
- **PCs (AMD 486 Clones):** Performance was generally comparable to Intel's 486 chips at similar clock speeds.[224]
- **PCs (PowerPC 601 @ 60-80MHz):** Used in early Power Macs. Estimates range from ~10 MFLOPS [198] to potentially higher based on SPECfp92 scores (~80-90 [35]), suggesting maybe 20-30 MFLOPS (DP) capability. We estimate an average of 25 MFLOPS (DP).
- **GPUs (e.g., S3 Trio64, ATI Mach64):** These were primarily 2D accelerators with limited, fixed-function 3D capabilities. Their general-purpose floating-point performance was negligible compared to contemporary CPUs.[214] They did not contribute meaningfully to the total FLOPs sold in the context of this analysis.

Total FLOPs Calculation (Approximate):

- Servers/Mainframes/Workstations: ~1.5M units * ~30 MFLOPS/unit (DP, weighted avg) ≈ 4.5 x 10^13 FLOPs (0.045 PFLOPS)
- PCs (x86): ~47M units * ~6 MFLOPS/unit (DP, weighted avg 486/Pentium) ≈ 2.8 x 10^14 FLOPs (0.28 PFLOPS)
- PCs (PowerPC): ~1M units * ~25 MFLOPS/unit (DP) ≈ 2.5 x 10^13 FLOPs (0.025 PFLOPS)
- GPUs: (Negligible contribution to total DP FLOPs)
- **Total Estimated FLOPs Sold in 1994:** ≈ 0.35 PFLOPS (or 350 TFLOPS)

Architectural Breakdown:

- **Server/Workstation:** Dominated by RISC architectures (SPARC, Alpha, MIPS, PA-RISC) in terms of new FLOPs sold, with legacy mainframes contributing less. ~0.045 PFLOPS total.
- **PC-x86:** Intel held a dominant market share, estimated around 80% [209], with AMD and other clone makers (like Cyrix) holding the rest.[211] Intel contributed ~0.22 PFLOPS, AMD/Compatibles ~0.06 PFLOPS.
- **PC-PowerPC:** Primarily Apple Macintosh. ~0.025 PFLOPS.
- **GPU:** Split among various vendors (S3, ATI, etc.). Negligible FLOPs contribution.

Table 2: Estimated 1994 FLOPs Sold by Architecture

Architecture Category	Est. Units Sold	Est. Avg. FLOPs/Unit (MFLOPS, DP)	Est. Total FLOPs (TFLOPS)	% of Total FLOPs

Server/Mainframe/ Workstation (RISC/Other)	~1,500,000	~30	~45	~12.9%
PC - x86 Intel (486/Pentium)	~37,600,000	~6	~226	~64.6%
PC - x86 AMD/Compatible (486)	~9,400,000	~6	~56	~16.0%
PC - PowerPC (PPC 601)	~1,000,000	~25	~25	~7.1%
GPU (All Types)	~48,000,000+	Negligible	Negligible	~0.0%
Total	**~49,500,000**	**N/A**	**~352**	**100.0%**

(Note: All figures are estimates based on available data and stated assumptions. DP = Double Precision. GPU FLOPs considered negligible for general-purpose computation this year.)
Analysis and Implications:
The 1994 data underscores the consolidation of the computing market around the PC, specifically the x86 architecture. The sheer volume of PC sales meant that even with lower per-unit FLOPs compared to high-end workstations, the aggregate floating-point capacity deployed via PCs dwarfed that of all other segments combined. This demonstrates the powerful effect of market standardization and volume manufacturing in driving the overall availability of computational resources, even if individual system performance varied widely.
Within the dominant x86 segment, the competitive dynamic between Intel and AMD was crucial.[209] While Intel held the lion's share of the market, AMD's presence as a viable alternative spurred innovation and likely contributed to faster performance improvements and price reductions across the board. This competition accelerated the FLOPs-per-dollar trend for mainstream computing, further solidifying the PC's central role.
Although GPUs in 1994 were primarily focused on rendering pixels rather than performing general-purpose floating-point calculations, their architectural foundation based on parallel processing units was established.[214] The ~48 million PCs sold represented a massive potential installed base for future parallel computing capabilities. While their FLOPs contribution was negligible in 1994, this year marks the quiet beginning of a hardware category that would eventually dominate the FLOPs landscape. The development of graphics APIs like OpenGL [214] also laid groundwork for future programmability.

6. 2004: The Era of Multi-Core, Mobile, and Maturing GPUs

Market Context:
By 2004, the computing landscape had diversified significantly. PCs remained central, characterized by intense competition between Intel's Pentium 4 (with Hyper-Threading) and

early dual-core concepts, and AMD's Athlon 64 and Opteron processors, which introduced 64-bit extensions to the x86 architecture.93 The server market was heavily dominated by x86 architectures, although RISC systems (like IBM POWER, Sun SPARC) and Intel's Itanium architecture held specialized high-end niches.232 The most dramatic growth occurred in mobile computing; smartphones, though still nascent by today's standards, and Personal Digital Assistants (PDAs) were rapidly proliferating, powered almost universally by ARM-based processors.95 Concurrently, GPUs had evolved into highly parallel processors, essential for the rapidly growing PC gaming market and beginning to attract attention for high-performance computing (HPC) tasks (early GPGPU). Key players were Nvidia with its GeForce 6 series and ATI (soon to be acquired by AMD) with its Radeon X800 series.101

Unit Sales Estimates:

- **Personal Computers (PCs):** Global shipments were robust, estimated between 175 million (IDC forecast June '04 [104]) and 187 million units (Gartner forecast Feb '04 [105]). Final IDC numbers put the year at 177.5 million units.[112] Dell and HP were the leading vendors.[91]
- **Servers:** Worldwide shipments reached 6.3 million units, with x86 servers comprising the vast majority (e.g., 1.6 million x86 servers shipped in Q4 alone).[89] RISC/Itanium systems represented a small fraction of unit volume.[232]
- **Smartphones:** Estimates vary, but Strategy Analytics placed 2004 sales at 17.5 million units.[115] Canalys data suggests higher numbers, possibly closer to 30-40 million based on quarterly figures.[113] Gartner reported ~674 million total mobile phones sold, but didn't break out smartphones specifically in the cited snippet.[95] We use an estimate of 30 million smartphones. Nokia was the leader, with Symbian OS dominant.[113] ARM architecture was universal.
- **PDAs:** Shipments were around 12.3 million units globally.[98] PalmOne (using Palm OS) and HP (using Windows CE/Mobile) were major players. ARM architecture was universal.
- **Discrete GPUs:** Estimating the attach rate for discrete GPUs in PCs is challenging. Reports from later years suggest rates around 32-36%.[109] Applying a ~35% attach rate to the ~177.5M PCs suggests approximately 62 million discrete GPUs were sold for PCs. Nvidia and ATI were the primary competitors, with market shares fluctuating quarterly but roughly comparable over the year.[101] Integrated graphics, primarily from Intel, equipped the remaining ~115 million PCs.

FLOPs per Unit Estimates:

Performance increased significantly compared to 1994, with multi-GHz clock speeds and architectural enhancements like SSE extensions in x86.

- **Servers (Intel Xeon "Nocona" 3.6 GHz):** Linpack benchmarks show ~1.8 GFLOPS (DP).[26] Theoretical peak SP performance using SSE2/SSE3 would be higher, perhaps ~25-30 GFLOPS SP (1 core * 3.6 GHz * 2 pipes * 4 SP ops/pipe).
- **Servers (AMD Opteron "SledgeHammer" 2.4 GHz):** Theoretical peak performance ~9.6 GFLOPS SP (1 core * 2.4 GHz * 2 pipes * 2 SP ops/pipe). Linpack performance likely lower, perhaps 1-2 GFLOPS DP.[11] We estimate an average server CPU (mix of Intel/AMD,

dual-socket common) at ~40 GFLOPS SP.

- **PCs (Intel Pentium 4 "Prescott" 3.0 GHz):** Theoretical peak ~24 GFLOPS SP (1 core * 3.0 GHz * 2 pipes * 4 SP ops/pipe with HT). Real-world benchmarks suggest lower effective performance.[227]
- **PCs (AMD Athlon 64 "Newcastle" 3200+ 2.2 GHz):** Theoretical peak ~8.8 GFLOPS SP (1 core * 2.2 GHz * 2 pipes * 2 SP ops/pipe).[229] We estimate an average PC CPU sold in 2004 at ~15 GFLOPS SP (weighted average across Intel/AMD models).
- **Mobile (ARM9/ARM11 based):** Early ARM cores with VFP (Vector Floating-Point) extensions, like ARM11, offered performance around 1.3-2 MFLOPS/MHz (SP).[248] With typical clock speeds of a few hundred MHz (e.g., up to 400 MHz), peak performance was likely below 1 GFLOPS SP. We estimate an average of 0.2 GFLOPS SP across all smartphones and PDAs sold.
- **GPUs (Nvidia GeForce 6800 Ultra):** Reported performance varies. One source cites 45 GFLOPS SP [236], while theoretical calculations based on shader count and clock speed could suggest over 100 GFLOPS SP.[237]
- **GPUs (ATI Radeon X800 XT):** One source cites 182 GFLOPS SP [239], while others suggest theoretical peaks over 100 GFLOPS SP.[238] We estimate a weighted average across all discrete GPUs sold (high-end, mid-range, low-end) at ~60 GFLOPS SP.
- **Integrated GPUs (e.g., Intel Extreme Graphics):** Performance significantly lower than discrete GPUs, likely in the single-digit GFLOPS SP range. We estimate an average of 2 GFLOPS SP.

Total FLOPs Calculation (Approximate):

- Servers (x86): ~6.0M units * 40 GFLOPS SP/unit ≈ 2.4 x 10^17 FLOPs (0.24 PFLOPS)
- Servers (RISC/Itanium/Other): ~0.3M units * ~50 GFLOPS SP/unit ≈ 1.5 x 10^16 FLOPs (0.015 PFLOPS)
- PCs (x86 CPU): ~177.5M units * 15 GFLOPS SP/unit ≈ 2.66 x 10^18 FLOPs (2.66 EFLOPS)
- Mobile (ARM - Smartphone + PDA): ~42.3M units * 0.2 GFLOPS SP/unit ≈ 8.5 x 10^15 FLOPs (0.008 EFLOPS) (negligible)
- Discrete GPUs: ~62M units * 60 GFLOPS SP/unit ≈ 3.72 x 10^18 FLOPs (3.72 EFLOPS)
- Integrated GPUs: ~115.5M units * 2 GFLOPS SP/unit ≈ 2.3 x 10^17 FLOPs (0.23 PFLOPS)
- **Total Estimated FLOPs Sold in 2004:** ≈ 6.87 EFLOPS (ExaFLOPS = 1018)

Architectural Breakdown:

- **Server-x86:** Intel held ~82% CPU market share [93], suggesting ~0.20 PFLOPS from Intel servers and ~0.04 PFLOPS from AMD servers.
- **Server-Other:** Mix of RISC (IBM POWER, Sun SPARC) and Itanium. ~0.015 PFLOPS.
- **PC-x86:** Intel share ~82%, AMD ~16%.[93] Intel CPUs ~2.18 EFLOPS, AMD CPUs ~0.43 EFLOPS.
- **Mobile-ARM:** ARM architecture universal. ~0.008 EFLOPS.
- **GPU-Nvidia Discrete:** Share ~46-50%.[102] ~1.7-1.8 EFLOPS.
- **GPU-ATI/AMD Discrete:** Share ~50-51%.[102] ~1.8-1.9 EFLOPS.
- **GPU-Integrated/Other:** Primarily Intel. ~0.23 EFLOPS.

Table 3: Estimated 2004 FLOPs Sold by Architecture

Architecture Category	Est. Units Sold	Est. Avg. FLOPs/Unit (GFLOPS, SP)	Est. Total FLOPs (PFLOPS)	% of Total FLOPs
Server - x86 Intel	~4,900,000	~40	~196	~2.9%
Server - x86 AMD	~1,100,000	~40	~44	~0.6%
Server - RISC/Itanium/Other	~300,000	~50	~15	~0.2%
PC - x86 Intel CPU	~145,500,000	~15	~2183	~31.8%
PC - x86 AMD CPU	~28,400,000	~15	~426	~6.2%
Mobile - ARM CPU/SoC	~42,300,000	~0.2	~8	~0.1%
GPU - Nvidia Discrete	~30,000,000	~60	~1800	~26.2%
GPU - ATI/AMD Discrete	~32,000,000	~60	~1920	~28.0%
GPU - Integrated/Other	~115,500,000	~2	~231	~3.4%
Total	**~400,000,000**	**N/A**	**~6823**	**100.0%**

(Note: All figures are estimates based on available data and stated assumptions. SP = Single Precision. Mobile FLOPs are highly approximate. Total units exceed sum of PC/Server/Mobile due to double-counting GPUs in PCs.)

Analysis and Implications:

The 2004 data reveals a pivotal moment where the sheer computational power delivered by discrete GPUs began to overshadow that of CPUs. Multiplying the average estimated FLOPs per discrete GPU (~60 GFLOPS SP) by the estimated units sold (~62 million) yields a total of ~3.7 EFLOPS. In contrast, the total FLOPs from all PC CPUs sold (~177.5 million units * ~15 GFLOPS SP/unit) amounts to ~2.7 EFLOPS. This signifies a fundamental shift: for the first time, the aggregate theoretical floating-point performance entering the market via discrete graphics cards exceeded the aggregate performance from the CPUs powering the vast majority of personal computers. This transition was primarily driven by the insatiable demands of 3D gaming, which pushed GPU vendors like Nvidia and ATI to develop highly parallel architectures capable of massive floating-point throughput.

Simultaneously, the mobile revolution, powered by ARM architecture, was establishing a colossal volume base.[99] While the individual FLOPs capability of smartphones and PDAs in 2004 was minimal (estimated <1 GFLOPS SP per device), the sale of tens of millions of units laid the foundation for ARM's future dominance in mobile and its eventual expansion into other computing segments. The focus on power efficiency inherent in ARM's design philosophy proved crucial for battery-powered devices and would later become increasingly

important across all computing domains.
Although GPUs were becoming the primary source of raw FLOPs, their application was still largely confined to graphics rendering. However, the programmability of GPUs was increasing (e.g., supporting DirectX 9.0 Shader Model 3.0 [240]), and researchers were beginning to explore their use for general-purpose computation (GPGPU) in scientific domains.[214] This nascent trend, leveraging the massive parallel floating-point power initially developed for gaming, set the stage for the later AI revolution, which would become heavily reliant on GPU-based acceleration. The server market, while critical for infrastructure, contributed a relatively small percentage of the total *new* FLOPs sold compared to the client market (PCs and GPUs combined), highlighting the scale difference driven by consumer electronics volumes.

7. 2014: The Mobile, Cloud, and Early AI Era

Market Context:
By 2014, the computing landscape was characterized by the ubiquity of mobile devices and the rise of cloud computing. Smartphone shipments, dominated by Apple's iOS and Google's Android operating systems running on ARM-based processors, vastly outnumbered PC shipments.123 Tablets, also primarily ARM-based, had carved out a significant market segment.125 The traditional PC market, while still substantial, faced pressure from these mobile form factors and experienced slower growth, though x86 architectures from Intel and AMD remained dominant.121 Cloud computing had become mainstream, fueling demand for large-scale data centers filled primarily with x86 servers.119 In the high-performance computing (HPC) and gaming segments, GPUs had become indispensable. Nvidia (Maxwell architecture) and AMD (GCN architecture) offered GPUs with TFLOPS-level performance, and their use in data centers for scientific computing and the early stages of the deep learning AI boom was accelerating.127
Unit Sales Estimates:

- **Personal Computers (PCs):** Global shipments were around 308-315 million units, showing stabilization after earlier declines.[121] Lenovo, HP, and Dell were the top vendors.
- **Servers:** Worldwide shipments reached approximately 9.2 million units, with x86 servers constituting the vast majority.[119]
- **Smartphones:** Global shipments surged to approximately 1.3 billion units.[123] Samsung and Apple led in market share, but competition from Chinese vendors like Huawei, Lenovo, and Xiaomi was intensifying. ARM architecture was universal in this segment.
- **Tablets:** Shipments totaled around 230 million units worldwide.[125] Apple's iPad maintained leadership, followed by Samsung and others, predominantly using ARM SoCs.
- **Discrete GPUs:** The attach rate for discrete GPUs in PCs hovered around 32-34%.[127] Based on ~310M PCs, this suggests roughly 100-105 million discrete GPUs were sold for client systems. Nvidia held a commanding market share lead over AMD, approximately 68% to 31%.[134] Integrated graphics (mostly Intel, some AMD APUs) equipped the remaining PCs. Data center GPU sales, while growing rapidly in importance for HPC/AI,

represented a much smaller unit volume compared to client GPUs at this time.

FLOPs per Unit Estimates:

Performance continued to climb, driven by multi-core CPUs, wider vector units (AVX2/FMA3 in x86), and increasingly powerful GPUs.

- **Servers (Intel Xeon E5 v3 "Haswell-EP"):** These processors featured up to 18 cores and AVX2/FMA3 instructions, enabling 16 double-precision (DP) or 32 single-precision (SP) FLOPs per core per cycle.[251] A mid-range 12-core chip at ~2.5 GHz could theoretically peak around 1.2 TFLOPS (DP) or 2.4 TFLOPS (SP) per CPU. A typical dual-socket server might average ~3-4 TFLOPS (SP).
- **Servers (AMD Opteron 6300 "Piledriver"):** Based on an older architecture with FMA4 support (8 DP or 16 SP FLOPs/cycle per module/2 cores).[253] A 16-core (8-module) chip at ~3.0 GHz might peak around 0.4 TFLOPS (DP) or 0.8 TFLOPS (SP) per CPU. Average server performance significantly lower than contemporary Intel Xeons, perhaps ~1 TFLOPS (SP) for a dual-socket system.
- **PCs (Intel Core i7 "Haswell" - e.g., i7-4790K):** With 4 cores, AVX2/FMA3, and turbo speeds up to 4.4 GHz, peak theoretical performance reached ~0.56 TFLOPS (DP) or ~1.1 TFLOPS (SP).[255] Average performance across the ~310M PCs sold (including lower-end Core i3/i5, Pentiums, Celerons) would be much lower, estimated at ~0.2 TFLOPS (SP).
- **PCs (AMD "Piledriver" / "Steamroller" - e.g., FX-8350):** Peak performance lower than Intel counterparts due to architecture.[257] Estimated average across AMD-based PCs ~0.1 TFLOPS (SP).
- **Mobile (High-end ARM SoCs - e.g., Apple A8, Qualcomm Snapdragon 805):** Integrated GPUs became the main FLOPs contributor. Apple A8's PowerVR GX6450 GPU was estimated around 136 GFLOPS (SP).[259] Qualcomm Adreno 420 in Snapdragon 805 likely offered comparable or slightly higher performance, perhaps ~200-300 GFLOPS (SP).[260] Average performance across the ~1.5 billion smartphones and tablets sold (including many lower-end devices) is estimated at ~0.1 TFLOPS (SP).
- **GPUs (Nvidia GeForce GTX 980 "Maxwell"):** Offered peak performance around 5 TFLOPS (SP).[262]
- **GPUs (AMD Radeon R9 290X "GCN"):** Offered peak performance around 5.6 TFLOPS (SP).[264] The weighted average across all ~100M discrete client GPUs sold (including mid-range and low-end cards) is estimated at ~2.5 TFLOPS (SP).
- **Integrated GPUs (Intel HD Graphics, AMD APUs):** Performance varied but significantly lower than discrete GPUs, perhaps averaging ~0.05 TFLOPS (SP).

Total FLOPs Calculation (Approximate):

- Servers (x86): ~9.0M units * ~3.5 TFLOPS SP/unit (weighted avg Intel/AMD) ≈ 3.15 x 10^19 FLOPs (31.5 EFLOPS)
- Servers (Other - POWER, SPARC): ~0.2M units * ~1.0 TFLOPS SP/unit ≈ 2.0 x 10^17 FLOPs (0.2 EFLOPS)
- PCs (x86 CPU): ~310M units * ~0.18 TFLOPS SP/unit (weighted avg Intel/AMD) ≈ 5.6 x 10^19 FLOPs (56 EFLOPS)
- Mobile (ARM - Smartphone + Tablet): ~1.53B units * ~0.1 TFLOPS SP/unit ≈ 1.53 x 10^20

FLOPs (153 EFLOPS)

- Discrete GPUs (Client + early Data Center): ~105M units * ~2.5 TFLOPS SP/unit ≈ 2.63 x 10^20 FLOPs (263 EFLOPS)
- Integrated GPUs: ~205M PCs/Servers * ~0.05 TFLOPS SP/unit ≈ 1.0 x 10^19 FLOPs (10 EFLOPS)
- **Total Estimated FLOPs Sold in 2014:** ≈ 514 EFLOPS

Architectural Breakdown:

- **Server-x86:** Intel dominated market share (~98%+). Intel ~31 EFLOPS, AMD ~0.5 EFLOPS.
- **Server-Other:** Primarily IBM POWER. ~0.2 EFLOPS.
- **PC-x86:** Intel share ~89%, AMD ~11%.[131] Intel CPUs ~50 EFLOPS, AMD CPUs ~6 EFLOPS.
- **Mobile/Client-ARM:** ARM architecture universal.[99] ~153 EFLOPS.
- **GPU-Nvidia Discrete:** ~68% unit share.[134] ~179 EFLOPS.
- **GPU-AMD Discrete:** ~31% unit share.[134] ~81 EFLOPS.
- **GPU-Integrated/Other:** Primarily Intel HD Graphics, AMD APUs. ~10 EFLOPS.

Table 4: Estimated 2014 FLOPs Sold by Architecture

Architecture Category	Est. Units Sold	Est. Avg. FLOPs/Unit (TFLOPS, SP)	Est. Total FLOPs (EFLOPS)	% of Total FLOPs
Server - x86 Intel	~8,800,000	~3.5	~30.8	~6.0%
Server - x86 AMD	~200,000	~1.0	~0.2	~0.0%
Server - Other (POWER, etc.)	~200,000	~1.0	~0.2	~0.0%
PC - x86 Intel CPU	~276,000,000	~0.18	~49.7	~9.7%
PC - x86 AMD CPU	~34,000,000	~0.1	~3.4	~0.7%
Mobile - ARM CPU/SoC	~1,530,000,000	~0.1	~153.0	~29.8%
GPU - Nvidia Discrete	~71,000,000	~2.5	~177.5	~34.5%
GPU - AMD Discrete	~32,000,000	~2.5	~80.0	~15.6%
GPU - Integrated/Other	~207,000,000	~0.05	~10.4	~2.0%
Total	**~2,160,000,000**	**N/A**	**~505.2**	**100.0%**

(Note: All figures are estimates based on available data and stated assumptions. SP = Single Precision. Server/PC CPU FLOPs represent CPU contribution only. Mobile SoC FLOPs dominated by integrated GPU. Total units exceed sum of PC/Server/Mobile/Tablet due to double-counting GPUs.)

Analysis and Implications:

The 2014 data shows a compute landscape dominated by client devices in terms of both unit volume and aggregate FLOPs. The sheer scale of the smartphone and tablet market, powered almost exclusively by ARM SoCs with increasingly capable integrated GPUs, meant that mobile devices likely contributed the largest single chunk of client-side FLOPs sold globally, potentially surpassing the contribution from PC CPUs.99 This underscores the massive shift in computing towards mobile platforms during this period.

However, discrete GPUs maintained their position as the primary source of *peak* floating-point performance entering the market overall. The combined FLOPs from Nvidia and AMD discrete GPUs (~260 EFLOPS) significantly exceeded the total estimated FLOPs from all CPUs combined (Servers ~31 EFLOPS + PCs ~53 EFLOPS + Mobile ~153 EFLOPS ≈ 237 EFLOPS). This reflects the continued divergence between highly parallel processors optimized for graphics (and increasingly, HPC/early AI) and general-purpose CPUs. Nvidia solidified its market share lead over AMD in the discrete GPU segment during this period.[134]

While data centers were becoming increasingly critical infrastructure, housing powerful multi-core x86 servers [251], their contribution to the *total annual volume* of FLOPs sold remained relatively small compared to the client market (PCs, mobile devices, and client GPUs). A server CPU, while powerful individually, was vastly outnumbered by the processors shipping in consumer devices. This highlights that the bulk of deployed computational power, measured by theoretical peak FLOPs, resided at the edge and in consumer hands, even as the strategic importance of centralized cloud computing grew. This period set the stage for the next major shift, where the computational demands of AI would dramatically elevate the importance and FLOPs contribution of data center hardware.

8. 2024: The AI Accelerator Revolution

Market Context:

The year 2024 is defined by the transformative impact of Artificial Intelligence, particularly generative AI. This has led to an unprecedented surge in demand for specialized compute hardware, primarily in data centers. GPUs and other AI accelerators, such as Google's TPUs and custom ASICs developed by hyperscalers, have become the dominant drivers of computational power growth, commanding massive investments.149 Nvidia, with its H100/H200 "Hopper" and newly announced "Blackwell" architectures, holds a commanding position in this lucrative market, with AMD's Instinct MI300 series providing the main competition.5 This AI boom has dramatically increased server market revenues, even if unit growth is more modest, due to the high average selling price (ASP) of AI-optimized servers.88 In the traditional server and PC CPU markets, Intel (with Emerald Rapids Xeons and Core Ultra client CPUs) and AMD (with EPYC Genoa/Turin server CPUs and Ryzen client CPUs) continue to compete fiercely within the x86 architecture.[141] However, ARM-based processors are making significant inroads. Cloud providers like AWS (Graviton) and Google are deploying custom ARM server chips, Nvidia offers its Grace CPU Superchip, and companies like Ampere Computing focus solely on ARM servers.[137] In the PC space, Apple's M-series silicon (ARM) continues its success, and Qualcomm's Snapdragon X Elite/Plus chips are powering a new wave of Windows-on-ARM laptops, challenging x86's historical dominance.[277] Furthermore, the open-standard RISC-V architecture is gaining momentum, particularly in embedded systems and IoT, with aspirations to enter the PC and server markets.[282]

The PC and smartphone markets, after pandemic-driven highs, experienced significant declines in 2023 but showed signs of stabilization and recovery in 2024.[139] Client GPUs, such as Nvidia's RTX 40/50 series and AMD's Radeon RX 7000/9000 series, continue to offer substantial performance for gaming and content creation, though unit shipments remain below peak levels seen in previous cycles.[147] The concept of the "AI PC," incorporating dedicated Neural Processing Units (NPUs), is also emerging.[281]

Unit Sales Estimates (2024):

- **Personal Computers (PCs):** Global shipments estimated around 263 million units, showing slight growth over 2023.[140] Lenovo, HP, and Dell remain the top vendors.
- **Servers:** Unit shipments likely around 12-14 million, but revenue growth is far more significant due to high-value AI servers.[88] x86 still leads in volume, but ARM server shipments are growing, particularly among hyperscalers.[137]
- **Smartphones:** Global shipments estimated around 1.24 billion units, showing modest growth.[144] Apple and Samsung lead, with strong growth from Transsion and Xiaomi. ARM architecture is universal.
- **Data Center GPUs/AI Accelerators:** Unit numbers are difficult to pinpoint precisely, as sales are often reported in revenue terms. However, given Nvidia's dominance [136] and the massive scale of deployments (e.g., hundreds of thousands of H100s by single companies [149]), total unit shipments of high-end accelerators are likely in the range of 3-5 million units, with a much higher ASP than client GPUs or CPUs.
- **Discrete Client GPUs:** Shipments estimated around 35 million units for the year.[147] Nvidia dominates market share (~88-90%), with AMD holding most of the remainder.[134]

FLOPs per Unit Estimates (Note Varying Precisions):

The FLOPs landscape in 2024 is dominated by AI accelerators operating at lower precisions.

- **Data Center GPUs (Nvidia H100 SXM):** Peak theoretical performance is staggering: ~67 TFLOPS (FP64 Tensor Core), ~989 TFLOPS (TF32 Tensor Core), ~1979 TFLOPS (FP16 Tensor Core), and ~3958 TFLOPS (FP8 Tensor Core).[6]
- **Data Center GPUs (AMD Instinct MI300X):** Peak theoretical performance: ~123 TFLOPS (FP64/FP32 Vector), ~980 TFLOPS (BF16), ~1961 TFLOPS (FP8).[7] Measured performance using optimized matrix multiplication (MAF) suggests achievable rates closer to ~654 TFLOPS (FP16), ~708 TFLOPS (BF16), and ~1273 TFLOPS (FP8).[8] For aggregation, using an average "effective AI FLOPs" value around 1500 TFLOPS (1.5 PFLOPS) per high-end accelerator seems plausible, acknowledging this blends different precisions.
- **Servers (Intel Xeon "Emerald Rapids"):** Peak theoretical performance per CPU around 10-20 TFLOPS (SP) using AVX-512.[273]
- **Servers (AMD EPYC "Genoa"):** Peak theoretical performance per CPU around 7-11 TFLOPS (SP) using AVX-512.[275] Average server CPU performance (mix of Intel/AMD, various core counts) estimated at ~15 TFLOPS (SP).

- **Servers (ARM - e.g., Ampere, Nvidia Grace):** Performance varies, but likely lower peak FLOPs per core than x86, compensated by core count and efficiency. Estimated average ~5 TFLOPS (SP) per chip.
- **PCs (High-end x86 - Core Ultra, Ryzen 7/9):** Peak performance around 1-3 TFLOPS (SP). Average across all PCs sold estimated at ~0.5 TFLOPS (SP).
- **PCs (ARM - Snapdragon X Elite, Apple M3):** Integrated GPU performance is significant. Snapdragon X Elite Adreno GPU peaks at ~4.6 TFLOPS (SP).[279] Apple M3 Max 40-core GPU estimated ~15-20 TFLOPS (SP)? (Extrapolating from M1's 2.6 TFLOPS [277]). Average across ARM PCs estimated at ~3 TFLOPS (SP).
- **Mobile (High-end ARM SoCs):** Integrated GPUs offer ~1-3 TFLOPS (SP). Average across all smartphones sold estimated at ~0.8 TFLOPS (SP).
- **Client GPUs (Nvidia RTX 4090):** Peak performance ~83 TFLOPS (SP).[286]
- **Client GPUs (AMD RX 7900 XTX):** Peak performance ~61 TFLOPS (SP).[288] Weighted average across all ~35M discrete client GPUs sold estimated at ~40 TFLOPS (SP).
- **Integrated GPUs (PC/Server):** Performance increasing but still far below discrete. Estimated average ~0.1 TFLOPS (SP).

Total FLOPs Calculation (Approximate - Dominated by Lower Precision AI FLOPs):

- Data Center Accelerators: ~4M units * 1500 TFLOPS/unit (Mixed AI Precision) ≈ 6 x 10^24 FLOPs (6 YFLOPS)
- Servers (CPUs): ~13M units * ~14 TFLOPS SP/unit (weighted x86/ARM) ≈ 1.8 x 10^20 FLOPs (0.18 ZFLOPS)
- PCs (CPU + iGPU): ~263M units * ~0.6 TFLOPS SP/unit (weighted x86/ARM + iGPU) ≈ 1.6 x 10^20 FLOPs (0.16 ZFLOPS)
- Mobile (ARM SoC): ~1.24B units * ~0.8 TFLOPS SP/unit ≈ 9.9 x 10^20 FLOPs (0.99 ZFLOPS)
- Discrete Client GPUs: ~35M units * ~40 TFLOPS SP/unit ≈ 1.4 x 10^21 FLOPs (1.4 ZFLOPS)
- Integrated GPUs (Standalone estimate): ~228M PCs * ~0.1 TFLOPS SP/unit ≈ 2.3 x 10^19 FLOPs (0.02 ZFLOPS)
- **Total Estimated FLOPs Sold in 2024:** ≈ 8.75 YFLOPS (YottaFLOPS = 1024)

(Note: This total is overwhelmingly dominated by the lower-precision FLOPs from AI accelerators. The contribution from traditional CPU/Client GPU FLOPs measured in SP/DP is several orders of magnitude smaller, likely in the range of 2-3 ZFLOPS total.)

Architectural Breakdown (Focus on FLOPs Drivers):

- **Data Center Accelerator:** Nvidia holds ~90%+ effective market share.[136] Nvidia ~5.4 YFLOPS, AMD/Other ~0.6 YFLOPS (primarily lower precision AI FLOPs).
- **Data Center CPU:** x86 still holds majority revenue/unit share, but ARM's share is growing rapidly (~15-20% units).[137] FLOPs contribution (~0.18 ZFLOPS SP) is negligible compared to accelerators.
- **Client CPU/SoC:** x86 dominates PCs (~75-80% share [131]), ARM dominates mobile [146] and is growing in PCs via Apple/Qualcomm.[281] Total FLOPs contribution (~1.15 ZFLOPS SP) is minor compared to accelerators.

- **Client GPU:** Nvidia dominates discrete share (~88-90%).[134] Total FLOPs contribution (~1.4 ZFLOPS SP) is minor compared to accelerators.

Table 5: Estimated 2024 FLOPs Sold by Dominant Architecture Categories

Architecture Category	Est. Units Sold (Millions)	Est. Avg. FLOPs/Unit (TFLOPS, Mixed Precision)	Est. Total FLOPs (YFLOPS)	% of Total FLOPs
Data Center Accelerator - Nvidia	~3.6	~1500 (AI Mix)	~5.40	~61.7%
Data Center Accelerator - AMD/Other	~0.4	~1500 (AI Mix)	~0.60	~6.9%
Mobile SoC (ARM)	~1240	~0.8 (SP)	~0.99	~11.3%
Discrete Client GPU (All)	~35	~40 (SP)	~1.40	~16.0%
PC CPU/SoC (x86 + ARM)	~263	~0.6 (SP)	~0.16	~1.8%
Server CPU (x86 + ARM)	~13	~14 (SP)	~0.18	~2.1%
Integrated Client GPU (non-SoC)	~228	~0.1 (SP)	~0.02	~0.2%
Total	**~1780**	**N/A**	**~8.75**	**100.0%**

(Note: All figures are highly estimated. AI Accelerator FLOPs dominate and are based on mixed/lower precision (FP16/FP8/TF32). Other categories primarily estimated using SP (FP32). Total units count overlaps categories.)

Analysis and Implications:

The 2024 data unequivocally demonstrates a paradigm shift in computing, driven by the demands of AI. The overwhelming majority (estimated >90% if considering only accelerator FLOPs vs all others combined, or ~69% based on the table above which includes mobile/client GPU SP FLOPs) of theoretical FLOPs sold now originate from specialized data center accelerators, primarily GPUs optimized for AI training and inference. This marks a fundamental change from previous eras where general-purpose CPUs in PCs or mobile devices contributed the largest share of aggregate FLOPs. The use of lower-precision number formats (FP16, BF16, FP8) in these accelerators is key to achieving the YottaFLOPS-scale throughput figures reported 5, making direct comparison with traditional FP32/FP64 CPU FLOPs misleading without careful qualification.

This shift has led to extreme market concentration. Nvidia's dominance in the AI accelerator space (~90% share) means it controls the vast majority of the new computational capacity being deployed globally.[136] This gives Nvidia significant pricing power and strategic influence,

although competition from AMD's Instinct line and custom silicon efforts by major cloud providers (like Google TPUs, AWS Trainium/Inferentia) is intensifying.[7]
While overshadowed in the total FLOPs count by AI accelerators, significant architectural diversification is occurring in the CPU market. ARM-based processors are mounting a serious challenge to x86's long-held dominance in both servers (driven by hyperscaler adoption for efficiency and custom designs [137]) and PCs (led by Apple's M-series and Qualcomm's Snapdragon X series for Windows [281]). Concurrently, the open-source RISC-V architecture is gaining traction, offering a royalty-free, customizable alternative that is attractive for specialized applications and potentially future mainstream adoption.[282] This suggests a future compute landscape that is far more heterogeneous than the largely x86-centric world of 2004 and 2014.

9. Looking Ahead: Projections for 2034 and 2044

Methodology:
Projecting computing FLOPs sold ten and twenty years into the future is inherently speculative. This forecast relies on extrapolating recent trends, particularly the growth rates observed in AI compute demand and semiconductor industry forecasts, tempered by qualitative assessments of potential drivers and constraints identified by industry experts and market analysts. Compound Annual Growth Rates (CAGRs) are used where available, but the immense uncertainty, especially for 2044, must be emphasized.

Growth Drivers:

- **Artificial Intelligence:** The demand for computation for both training increasingly large AI models and deploying them for inference across myriad applications is expected to remain the primary driver of FLOPs growth.[266] While current growth rates in training compute (doubling every 5-6 months [271]) are likely unsustainable long-term [293], the overall demand for AI-related FLOPs is projected to grow substantially, potentially at CAGRs of 20-35% or higher through the next decade.[268] The shift towards more complex reasoning and agentic AI systems could further fuel inference compute demand.[290]
- **Architectural Specialization and Heterogeneity:** The trend towards specialized processors (GPUs, TPUs, NPUs, potentially quantum co-processors) tailored for specific workloads (AI, graphics, scientific simulation) will likely continue.[291] This allows for greater efficiency (FLOPS/Watt) compared to general-purpose CPUs. The co-existence of multiple CPU architectures (x86, ARM, RISC-V) catering to different market segments (performance, efficiency, customization, cost) will likely persist and intensify.[281]
- **Edge Computing:** Processing data closer to its source for applications like autonomous vehicles, IoT analytics, and real-time control systems will drive demand for compute power outside traditional data centers. Global spending on edge computing is forecast to grow at nearly 14% CAGR, reaching $380 billion by 2028.[297]
- **Continued Semiconductor Advancement:** While Dennard scaling has ended and Moore's Law is slowing for monolithic chips, advancements in materials, transistor

designs (like Gate-All-Around FETs), and especially advanced packaging techniques (chiplets, heterogeneous integration) continue to enable performance and density improvements.[269]

Potential Constraints:

- **Power Consumption:** This is arguably the most significant constraint. Data centers already consume a substantial fraction of global electricity, and AI workloads are particularly power-hungry.[266] Projections suggest AI power demand could exceed the current consumption of entire countries by 2030.[267] Sustaining exponential FLOPs growth may require radical improvements in energy efficiency (FLOPS/Watt) or massive investments in energy generation and infrastructure, potentially limiting deployment scale.[269]
- **Economic Factors:** The cost of building leading-edge semiconductor fabs runs into the tens of billions of dollars. The capital expenditure required for large-scale AI supercomputers is also immense (e.g., hundreds of billions projected for future systems [272]). Economic downturns, shifts in investment priorities (e.g., if AI monetization lags [270]), or rising costs could dampen the growth rate.
- **Software and Algorithmic Efficiency:** Effectively utilizing hardware capable of YottaFLOPS or ZettaFLOPS (1027) requires significant advances in parallel programming models, algorithms, and software tools. AI model efficiency is also crucial; if models become significantly more efficient, the demand for raw FLOPs might grow more slowly.
- **Geopolitics and Supply Chains:** Increasing geopolitical tensions, trade restrictions (especially concerning advanced semiconductors [290]), and supply chain vulnerabilities pose risks to the global manufacturing and deployment of cutting-edge compute hardware.[270]
- **Physical Limits:** While advanced packaging provides avenues for continued performance scaling, fundamental physical limits related to heat dissipation, interconnect bandwidth, and ultimately quantum effects will eventually constrain computational density and speed.[198]

2034 Projection:

- **Total FLOPs:** Continued strong, double-digit CAGR (perhaps 20-30% average [268]) driven primarily by AI accelerator deployments seems plausible. This would place total annual FLOPs sold in the range of **100-500 YFLOPS**, potentially reaching the low ZettaFLOPS scale (1027 FLOPs). The precision basis will remain mixed, dominated by formats suitable for AI.
- **Architectural Mix:**
    - AI Accelerators (GPUs, ASICs, etc.): Will constitute the vast majority (>95%) of total FLOPs sold. Nvidia is likely to remain dominant, but competition from AMD, Intel, and custom silicon from hyperscalers will intensify.
    - Server CPUs: ARM architectures are expected to gain significant share, potentially exceeding 50% of the data center CPU market by the late 2020s or early 2030s, driven by hyperscaler adoption and efficiency demands.[137] x86 will likely retain a substantial share, particularly in enterprise. RISC-V may start appearing in niche

server roles.[282] Overall CPU FLOPs contribution remains minor compared to accelerators.
    - Client CPUs/SoCs: A mix of x86 and ARM is expected in PCs.[281] ARM will continue to dominate smartphones and tablets. RISC-V may gain share in lower-end mobile/IoT and potentially PCs.[284] Aggregate FLOPs from client devices will be significant due to volume but dwarfed by data center accelerators.

2044 Projection (Highly Speculative):

- **Total FLOPs:** Extrapolating another decade is fraught with uncertainty. Assuming growth moderates significantly due to constraints (e.g., averaging 10-15% CAGR from 2034-2044), total FLOPs sold could reach the scale of **1-10 ZettaFLOPS** (1027–1028 FLOPs). Reaching this scale likely depends heavily on breakthroughs in energy efficiency and potentially new computing paradigms.
- **Architectural Mix:**
    - AI/Specialized Accelerators: Will still dominate. The specific technologies may evolve (e.g., more ASICs, potential integration of photonic or neuromorphic elements, early quantum co-processors [291]).
    - General-Purpose CPUs: The balance between x86, ARM, and RISC-V is difficult to predict. If energy efficiency and customization become paramount, ARM and RISC-V could marginalize x86 significantly.[282] However, the incumbency and software ecosystem of x86 could ensure its persistence.
    - New Paradigms: It's possible that fundamentally different computing approaches (quantum, optical, biological) could start contributing measurable FLOPs, although widespread deployment by 2044 remains uncertain.[198]

Table 6: Projected 2034 & 2044 FLOPs Sold by Dominant Architecture Categories

Year	Architecture Category	Estimated Total FLOPs (Mixed Precision)	Projected CAGR	Key Assumptions/Uncertainties
2034	AI Accelerators (DC)	100 - 500 YFLOPS (Low ZFLOPS)	20-30% (2024-2034)	Continued strong AI investment, advances in accelerator efficiency, manageable power constraints. High uncertainty in growth rate.
2034	General Purpose CPUs (All)	Low ZFLOPS	<15% (2024-2034)	x86 vs ARM vs RISC-V share shift, dwarfed by accelerators.
2034	Client GPUs/SoCs (All)	Low ZFLOPS	<15% (2024-2034)	Driven by mobile/PC volume,

				dwarfed by accelerators.
2034 Total	**All**	**~100 - 500 YFLOPS+**	**~20-30% (Overall)**	**Dominated by AI accelerators; overall rate depends heavily on AI demand and efficiency gains.**
2044	AI Accelerators (DC)	1 - 10 ZFLOPS	10-15% (2034-2044)	Moderating AI growth, significant efficiency breakthroughs, potential new accelerator types, power/cost constraints bite.
2044	General Purpose CPUs (All)	Mid ZFLOPS?	<10% (2034-2044)	Architectural balance highly uncertain (x86/ARM/RISC-V).
2044	Client GPUs/SoCs (All)	Mid ZFLOPS?	<10% (2034-2044)	Volume driven, potential integration with new paradigms.
2044	New Paradigms (Quantum etc.)	Highly Uncertain (Low ZFLOPS?)	N/A	Dependent on major technological breakthroughs.
2044 Total	**All**	**~1 - 10 ZFLOPS+**	**~10-15% (Overall)**	**Highly speculative; assumes continued innovation overcomes major power/cost/physical constraints.**

(Note: FLOPs figures are order-of-magnitude estimates, primarily reflecting lower-precision AI compute. CAGR figures are illustrative averages over the period.)

10. Conclusion and Strategic Insights

The history of computing over the past four decades is, in large part, a story of the relentless pursuit and deployment of floating-point computational power. From the sub-PetaFLOPS scale of 1984, dominated by mainframes and minicomputers, the annual volume of theoretical FLOPs sold globally has surged by an almost incomprehensible factor, reaching the YottaFLOPS (1024) range in 2024. This exponential growth has been fueled by successive architectural waves: the rise of the x86 PC which democratized computing and vastly increased unit volumes; the mobile revolution driven by power-efficient ARM SoCs scaling to billions of devices; and most recently, the explosion of parallel processing power delivered by GPUs and specialized AI accelerators.

The current era represents a fundamental inflection point. Since roughly the mid-2010s, and accelerating dramatically in the 2020s, the vast majority of *new* theoretical FLOPs entering the market originate not from general-purpose CPUs, but from data center accelerators designed specifically for AI workloads. Furthermore, these accelerators achieve their headline performance figures often using lower-precision arithmetic (FP16, BF16, FP8), optimized for the statistical nature of deep learning rather than the traditional FP32/FP64 precision used in scientific computing and earlier benchmarks. This makes direct FLOPs comparisons across eras and device classes complex, but underscores the degree to which AI has reshaped the hardware landscape. Nvidia's current dominance in this critical AI accelerator segment gives it unprecedented market influence.

While AI accelerators drive the headline FLOPs numbers, the underlying architectural landscape is becoming more diverse. The long-standing duopoly of x86 in PCs and servers is facing credible challenges. ARM's architecture, honed in the power-constrained mobile market, is proving viable and efficient for laptops (Apple Silicon, Windows on ARM) and increasingly for servers, particularly in large cloud deployments where power efficiency and customization are key. The open-standard RISC-V architecture also presents a long-term alternative, offering freedom from licensing fees and enabling greater customization, which is attracting significant interest across various segments from embedded systems to potentially HPC and data centers. The future of computing is likely to be increasingly heterogeneous, with different architectures optimized for different tasks and market segments.

However, the continuation of exponential growth in FLOPs deployment faces significant hurdles. The most pressing is energy consumption. The power required to operate large-scale AI training clusters and data centers is substantial and growing rapidly, potentially straining electricity grids and conflicting with climate goals.[266] Future progress will be heavily dependent on improving energy efficiency (FLOPS per Watt). The immense capital cost of building leading-edge semiconductor fabrication plants and deploying massive AI infrastructure also presents economic challenges.[270] Geopolitical factors impacting semiconductor supply chains add another layer of complexity and risk.[270]

These trends carry significant strategic implications:

- **Hardware Vendors:** Success will require continuous innovation in both performance and energy efficiency. Specialization for key workloads, particularly AI, is crucial.

Vendors must navigate a complex landscape with competing architectures (x86, ARM, RISC-V) and the rise of custom silicon designed by large customers (hyperscalers). Advanced packaging and heterogeneous integration will be key technology drivers.[269]

- **Software Developers:** The challenge lies in creating software that can effectively and efficiently utilize massively parallel, heterogeneous hardware. Optimizing algorithms for specific accelerators and leveraging lower-precision arithmetic where appropriate will be critical for performance. Developing more compute-efficient AI models is also essential.
- **Enterprises and Cloud Providers:** Strategic decisions regarding infrastructure investment are paramount. Balancing the performance benefits of cutting-edge hardware (especially AI accelerators) against cost and power consumption is critical. The choice between deploying proprietary hardware, utilizing cloud services, or developing custom silicon will depend on scale, workload, and strategic goals.[290]
- **Policymakers:** Governments face the challenge of fostering domestic semiconductor capabilities amidst geopolitical competition, ensuring stable supply chains, managing the energy demands of large-scale computing, and potentially regulating the development and deployment of powerful AI systems.

In conclusion, the journey from KiloFLOPS to YottaFLOPS sold annually within four decades represents one of the most profound technological transformations in history. While the exponential trajectory faces undeniable physical and economic constraints, the demand for computation, currently catalyzed by AI, shows few signs of abating. Future progress will likely depend not just on raw FLOPs increases, but on achieving those gains with greater efficiency and through increasingly diverse and specialized computing architectures.

Works cited

1. indianexpress.com, accessed May 7, 2025, https://indianexpress.com/article/explained/everyday-explainers/what-are-petaflops-in-computing-explained-8628459/#:~:text=FLOPs%20or%20Floating%2DPoint%20Operations,and%20artificial%20intelligence%20(AI).
2. The Importance of FLOPS and its Impact on Your PC's Speed and Efficiency - Lenovo, accessed May 7, 2025, https://www.lenovo.com/us/en/glossary/flops/
3. What Is FLOP? Floating-point Operations - Acronyms - Martech Zone, accessed May 7, 2025, https://martech.zone/acronym/flop/
4. Global computing capacity – AI Impacts, accessed May 7, 2025, https://aiimpacts.org/global-computing-capacity/
5. Nvidia cuts FP8 training performance in half on RTX 40 and 50 series GPUs - Reddit, accessed May 7, 2025, https://www.reddit.com/r/LocalLLaMA/comments/1ideaxu/nvidia_cuts_fp8_training_performance_in_half_on/
6. Choosing between NVIDIA H100 vs A100 - Performance and Costs Considerations, accessed May 7, 2025, https://blog.ori.co/choosing-between-nvidia-h100-vs-a100-performance-and-costs-considerations

7. AMD Instinct™ MI300 Series Platform | Solution - GIGABYTE Global, accessed May 7, 2025, https://www.gigabyte.com/Industry-Solutions/amd-instinct-mi300
8. Measuring Max-Achievable FLOPs – Part 2 - ROCm Blogs - AMD, accessed May 7, 2025, https://rocm.blogs.amd.com/software-tools-optimization/measuring-max-achievable-flops-part2/README.html
9. cowles.yale.edu, accessed May 7, 2025, https://cowles.yale.edu/sites/default/files/2022-08/d1324.pdf
10. Two Centuries of Productivity Growth in Computing | The Journal of Economic History, accessed May 7, 2025, https://www.cambridge.org/core/journals/journal-of-economic-history/article/two-centuries-of-productivity-growth-in-computing/856EC5947A5857296D3328FA154BA3A3
11. LINPACK benchmarks - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/LINPACK_benchmarks
12. Linpack Benchmark Results - Roy Longbottom's PC benchmark Collection, accessed May 7, 2025, http://www.roylongbottom.org.uk/linpack%20results.htm
13. Performance of Various Computers Using Standard Linear Equations Software, accessed May 7, 2025, https://www.researchgate.net/publication/2543892_Performance_of_Various_Computers_Using_Standard_Linear_Equations_Software
14. ES 9000-720 - TOP500, accessed May 7, 2025, https://top500.org/system/167149/
15. SPARC - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/SPARC
16. Linpack benchmarks in C for Sun SPARC systems - Math.Utah.Edu, accessed May 7, 2025, https://www.math.utah.edu/~beebe/java/c/Linpack-c.html
17. Benchmarks-LINPACK and MATLAB - Fame and fortune from megaflops by Cleve Moler - MathWorks, accessed May 7, 2025, https://ww2.mathworks.cn/content/dam/mathworks/tag-team/Objects/l/72911_92029v00Cleve_Benchmarks_LINPACK_MATLAB_Sum_Fall_1994.pdf
18. Past, Present, and Future* 1 LINPACK Benchmark History - Netlib.org, accessed May 7, 2025, https://www.netlib.org/utk/people/JackDongarra/PAPERS/hpl.pdf
19. Java Linpack Benchmark -- list of timings - The Netlib, accessed May 7, 2025, https://www.netlib.org/benchmark/linpackjava/rawTimings/all.html
20. Linpack Benchmark Results On PCs and Later Devices Roy Longbottom - ResearchGate, accessed May 7, 2025, https://www.researchgate.net/publication/319176419_Linpack_Benchmark_Results_On_PCs_and_Later_Devices_Roy_Longbottom
21. LINPACK and MATLAB Benchmarks - MathWorks, accessed May 7, 2025, https://www.mathworks.com/company/technical-articles/linpack-and-matlab-benchmarks.html
22. Intel Clock-Doubler 486 Debuts as 486DX2: 3/4/92 - CECS, accessed May 7, 2025, https://www.cecs.uci.edu/~papers/mpr/MPR/ARTICLES/060304.PDF
23. November 2004 - TOP500, accessed May 7, 2025, https://top500.org/lists/top500/2004/11/

24. Performance Comparison of Java/.NET Runtimes (Oct 2004), accessed May 7, 2025, http://www.shudo.net/jit/perf/
25. Linpack HPL Performance on IBM eServer 326 and xSeries 336 Servers, accessed May 7, 2025, https://public.dhe.ibm.com/eserver/benchmarks/wp_Linpack_072905.pdf
26. Frequent Asked Questions on the LINPACK Benchmark - Netlib.org, accessed May 7, 2025, https://www.netlib.org/utk/people/JackDongarra/faq-linpack.html
27. Opteron - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Opteron
28. AMD's Sledgehammer: By the Numbers - Datamation, accessed May 7, 2025, https://www.datamation.com/erp/amds-sledgehammer-by-the-numbers/
29. Java Linpack List of timings - Netlib.org, accessed May 7, 2025, https://www.netlib.org/benchmark/linpackjava/timings_list.html
30. Intel's Pentium 4 3.4GHz Processors Reviewed - Slashdot, accessed May 7, 2025, https://hardware.slashdot.org/story/04/03/22/0723247/intels-pentium-4-34ghz-processors-reviewed
31. AMD Athlon 64 3200+ Benchmarks - Geekbench, accessed May 7, 2025, https://browser.geekbench.com/processors/amd-athlon-64-3200
32. AMD Athlon 64 3200+ - CPU Benchmarks, accessed May 7, 2025, https://www.cpubenchmark.net/cpu.php?cpu=AMD+Athlon+64+3200%2B&id=67
33. Performance Analysis of Computer Systems - TU Dresden, accessed May 7, 2025, https://tu-dresden.de/zih/ressourcen/dateien/lehre/ws0910/lars/vorlesung_04_top500_stream_hpcc.pdf
34. CPU Specs Database - TechPowerUp, accessed May 7, 2025, https://www.techpowerup.com/cpu-specs/?f=codename_=SledgeHammer
35. Cost of CPU Performance Through Time 1944-2003 - jcmit.net, accessed May 7, 2025, https://jcmit.net/cpu-performance.htm
36. QG_GYRE and Memory Bandwidth, accessed May 7, 2025, http://www.cs.virginia.edu/~mccalpin/papers/balance/qgbox/case_study.html
37. Trends in Peak MFLOPS, SPECfp92, Sustainable Memory Bandwidth... - ResearchGate, accessed May 7, 2025, https://www.researchgate.net/figure/Trends-in-Peak-MFLOPS-SPECfp92-Sustainable-Memory-Bandwidth-Mwords-s-and_fig2_51992086
38. SPEC CPU 2000 Results - SPEC.org, accessed May 7, 2025, https://www.spec.org/cpu2000/results/
39. Table 1, accessed May 7, 2025, http://users.utcluj.ro/~baruch/book_ssce/SSCE-Benchmark.pdf
40. FLOPS – Knowledge and References - Taylor & Francis, accessed May 7, 2025, https://taylorandfrancis.com/knowledge/Engineering_and_technology/Computer_science/FLOPS/
41. Whetstone (benchmark) - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Whetstone_(benchmark)
42. Instructions per second - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Instructions_per_second
43. Full text of "Computerworld" - Internet Archive, accessed May 7, 2025, https://archive.org/stream/computerworld1740unse_1/computerworld1740unse_1

_djvu.txt
44. Full text of "Computerworld" - Internet Archive, accessed May 7, 2025, https://archive.org/stream/computerworld158unse/computerworld158unse_djvu.txt
45. Floating point performance of classic minicomputers ..., accessed May 7, 2025, https://retrocomputing.stackexchange.com/questions/4725/floating-point-performance-of-classic-minicomputers
46. DG Eagle MV/8000 / The Soul of a New Machine - Clemson University, accessed May 7, 2025, https://people.computing.clemson.edu/~mark/330/eagle.html
47. An overview of common benchmarks, accessed May 7, 2025, https://users.ece.utexas.edu/~ljohn/teaching/382m-15/reading/weicker.pdf
48. How to properly calculate CPU and GPU FLOPS performance?, accessed May 7, 2025, https://scicomp.stackexchange.com/questions/36306/how-to-properly-calculate-cpu-and-gpu-flops-performance
49. Calculating Double FLOPS for GPU - graphics card - Super User, accessed May 7, 2025, https://superuser.com/questions/781262/calculating-double-flops-for-gpu
50. Trends in the cost of computing - AI Impacts Wiki, accessed May 7, 2025, https://wiki.aiimpacts.org/ai_timelines/trends_in_the_cost_of_computing
51. Who Gains from Innovation? | NBER, accessed May 7, 2025, https://www.nber.org/digest/oct04/who-gains-innovation
52. www.nber.org, accessed May 7, 2025, https://www.nber.org/system/files/chapters/c8314/c8314.pdf
53. Apple Has Always Been a Niche Player | Low End Mac, accessed May 7, 2025, https://lowendmac.com/2003/apple-has-always-been-a-niche-player/
54. Market share of personal computer vendors - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Market_share_of_personal_computer_vendors
55. Worldwide PC Market Grew 24 Percent in 1995 - Tech Insider, accessed May 7, 2025, https://www.tech-insider.org/statistics/research/1996/0126.html
56. Intel Corporation Annual Report 1984, accessed May 7, 2025, https://www.intel.cn/content/dam/doc/report/history-1984-annual-report.pdf
57. The Personal Computer Market by Metropolitan Statistical Areas. - Geographicus Rare Antique Maps, accessed May 7, 2025, https://www.geographicus.com/P/AntiqueMap/personalcomputer-infoworld-1984
58. The Golden Age of InfoWorld Covers, 1984-1985 - Technologizer, accessed May 7, 2025, https://technologizer.com/2010/02/03/infoworld-covers/index.html
59. 1984 January Byte Magazine Computer Future Trends 1984 And Beyond - eBay, accessed May 7, 2025, https://www.ebay.com/itm/356560059036
60. Byte (magazine) - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Byte_(magazine)
61. Computerworld 1985-10-21: Vol 19 Iss 42 : Free Download, Borrow, and Streaming, accessed May 7, 2025, https://archive.org/details/sim_computerworld_1985-10-21_19_42/sim_computerworld_1985-10-21_19_42
62. Computer manufacturing: change and competition - Bureau of Labor Statistics,

accessed May 7, 2025, https://www.bls.gov/opub/mlr/1996/08/art3full.pdf
63. 19850601.pdf - Bitsavers.org, accessed May 7, 2025, http://bitsavers.org/magazines/Datamation/19850601.pdf
64. Fit at 40: the revolutionary Apple Mac in numbers - The Economic Times, accessed May 7, 2025, https://m.economictimes.com/tech/technology/fit-at-40-the-revolutionary-apple-mac-in-numbers/articleshow/107047120.cms
65. National Macintosh Computer Day: A Brief History - AGiRepair, accessed May 7, 2025, https://agirepair.com/national-macintosh-computer-day-a-brief-history/
66. Dog-Eat-Dog Shake-Out | TIME, accessed May 7, 2025, https://time.com/archive/6704275/dog-eat-dog-shake-out/
67. The Mainframe Era Is Fading -- And The Micro Is Taking Command - Tech Insider, accessed May 7, 2025, https://www.tech-insider.org/mainframes/research/1987/1130.html
68. Chapter V. Mass Computerization - IT Museum DataArt, accessed May 7, 2025, https://museum.dataart.com/history/chapter-5-mass-computerization
69. www.justice.gov, accessed May 7, 2025, https://www.justice.gov/atr/case-document/file/492361/dl
70. MS Annual Report - Management's Discussion and Analysis - Microsoft, accessed May 7, 2025, https://www.microsoft.com/investor/reports/ar96/financials/md.htm
71. Sun retains lion's share of workstation market in 1995 - SunWorld - January 1996, accessed May 7, 2025, http://sunsite.uakom.sk/sunworldonline/swol-01-1996/swol-01-workstation.html
72. The U.S. Personal Computer Industry, 1974-1994 - ResearchGate, accessed May 7, 2025, https://www.researchgate.net/figure/The-US-Personal-Computer-Industry-1974-1994_fig1_272306584
73. Computer Almanac - Numbers About Computers, accessed May 7, 2025, https://www.cs.cmu.edu/~bam/numbers.html
74. Desk Reference (INFORMATION PLEASE BUSINESS ALMANAC AND SOURCEBOOK) - Amazon.com, accessed May 7, 2025, https://www.amazon.com/Information-Business-Reference-INFORMATION-SOURCEBOOK/dp/0395643848
75. IDC Report Establishes Novell As World TCP/IP Leader | Micro Focus, accessed May 7, 2025, https://www.novell.com/news/press/archive/1995/08/pr00191.html
76. Worst PC sales drop in history, accessed May 7, 2025, https://money.cnn.com/2013/04/10/technology/pc-sales/
77. Flash Memory; Technology Alert, February 1994 - eGrove, accessed May 7, 2025, https://egrove.olemiss.edu/cgi/viewcontent.cgi?article=2849&context=aicpa_news
78. Dataquest predicts...1994 and beyond, 1994 - Computer History Museum - Archive Server, accessed May 7, 2025, http://archive.computerhistory.org/resources/access/text/2013/04/102723275-05-01-acc.pdf
79. 1994 - report - IBM, accessed May 7, 2025,

https://www.ibm.com/investor/att/pdf/IBM_Annual_Report_1994.pdf
80. CTIMES Article - Worldwide Server Shipments Grew 1.9%, Revenue Decreased 2.1%, accessed May 7, 2025, https://en.ctimes.com.tw/DispArt.asp?O=HJY148XW17EARASTDC
81. Usage share of operating systems - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Usage_share_of_operating_systems
82. securities and exchange commission - Investor Relations :: Intel Corporation (INTC), accessed May 7, 2025, https://www.intc.com/filings-reports/all-sec-filings/content/0000050863-95-000001/0000050863-95-000001.pdf
83. Shift to Pentium Begins in Earnest - CECS, accessed May 7, 2025, https://www.cecs.uci.edu/~papers/mpr/MPR/ARTICLES/090101.pdf
84. Power Macintosh - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Power_Macintosh
85. Solved Between 1991 and 1994, Apple Computer engaged in a | Chegg.com, accessed May 7, 2025, https://www.chegg.com/homework-help/questions-and-answers/1991-1994-apple-computer-engaged-holding-action-desktop-market-dominated-pcs-using-intel-c-q13743573
86. Industry News Bullets, accessed May 7, 2025, https://www.zeigassocs.com/newsbull.htm
87. IDC Says Serverware Market Will Near $6.4 Billion - HPCwire, accessed May 7, 2025, https://www.hpcwire.com/2000/08/11/idc-says-serverware-market-will-near-6-4-billion/
88. Servers Market Insights - IDC, accessed May 7, 2025, https://www.idc.com/promo/servers/
89. Q4 Server Sales Up 5.1% - Circuits Assembly Online Magazine, accessed May 7, 2025, https://circuitsassembly.com/ca/products-itemid-fix/290-news/2005-news/10264-q4-server-sales-up-51.html
90. Worldwide Server Market Hits $49 Billion - HPCwire, accessed May 7, 2025, https://www.hpcwire.com/2005/02/25/worldwide_server_market_hits_49_billion/
91. Gartner: U.S. PC Shipments Up 8.3% In 2004 | CRN, accessed May 7, 2025, https://www.crn.com/news/components-peripherals/57702068/gartner-u-s-pc-shipments-up-8-3-in-2004
92. Worldwide Handheld Shipments Gain Slightly - PalmInfocenter, accessed May 7, 2025, https://www.palminfocenter.com/news/6997/worldwide-handheld-shipments-gain-slightly/
93. CPU market shares for Q4 2004: Intel - 82.2%, AMD - 16.6% - ZDNet, accessed May 7, 2025, https://www.zdnet.com/article/cpu-market-shares-for-q4-2004-intel-82-2-amd-16-6/
94. CPU market shares in Q3 2005: Intel - 80.8%, AMD - 17.8% - ZDNET, accessed

May 7, 2025, https://www.zdnet.com/article/cpu-market-shares-in-q3-2005-intel-80-8-amd-17-8/
95. Gartner Says Strong Fourth Quarter Sales Led Worldwide Mobile Phone Sales to 30 Percent Growth in 2004 - Tech Insider, accessed May 7, 2025, https://www.tech-insider.org/mobile/research/2005/0305.html
96. UMTS and 3G market share distribution, accessed May 7, 2025, https://www.umtsworld.com/industry/user_equipment.htm
97. $4.3 bln of PDAs sold in 2004, average selling price was $353 - ZDNET, accessed May 7, 2025, https://www.zdnet.com/home-and-office/networking/4-3-bln-of-pdas-sold-in-2004-average-selling-price-was-353-6008000490/
98. Gartner: Worldwide PDA Shipments Grew 7% in 2004 - PalmInfocenter, accessed May 7, 2025, http://www.palminfocenter.com/news/7613/gartner-worldwide-pda-shipments-grew-7-in-2004/
99. Shaping the Connected World - Arm, accessed May 7, 2025, https://www.arm.com/-/media/arm-com/company/Legacy%20Financial%20PDFs/ARM_AR14_Strategic%20Report%20Final.pdf?la=en
100. ARM Holdings plc Preliminary Results For The Year Ended 31 December 2004, accessed May 7, 2025, https://www.design-reuse.com/news/9549/arm-holdings-plc-preliminary-year-ended-31-december-2004.html
101. ATI lost graphics market share to Nvidia in Q3 - The Register, accessed May 7, 2025, https://www.theregister.com/2006/12/06/q3_06_graphics_market/
102. Nvidia chisels away at ATI market share - The Register, accessed May 7, 2025, https://www.theregister.com/2005/01/26/graphics_chip_sales_q4_04/
103. AMD and NVIDIA AIB GPU Market Share from 2002 to 2016 - www.guru3d.com, accessed May 7, 2025, https://www.guru3d.com/story/amd-and-nvidia-aib-gpu-market-share-from-2002-to-2016/
104. IDC raises PC shipment forecast for 2004 - EE Times, accessed May 7, 2025, https://www.eetimes.com/idc-raises-pc-shipment-forecast-for-2004/
105. PC shipments to rise in 2004, says Gartner - CNET, accessed May 7, 2025, https://www.cnet.com/tech/computing/pc-shipments-to-rise-in-2004-says-gartner/
106. Smartphones: iPhone and the Big Fat Mobile Industry - RoughlyDrafted, accessed May 7, 2025, http://www.roughlydrafted.com/RD/RDM.Tech.Q1.07/BEC05CE1-D5EB-4E48-B46C-7385D5AADCFE.html
107. Market Watch – a report series on the Graphics Processor Unit market - Jon Peddie Research, accessed May 7, 2025, https://www.jonpeddie.com/store/market-watch/
108. JPR: 251 Million GPUs Shipped Globally in 2024, More Units than CPUs - TechPowerUp, accessed May 7, 2025,

https://www.techpowerup.com/331172/jpr-251-million-gpus-shipped-globally-in-2024-more-units-than-cpus
109. Nvidia, and AMD Increase GPU Attach Rates While Total GPU Shipments Remain Flat Quarter to Quarter - Design And Reuse, accessed May 7, 2025, https://www.design-reuse.com/news/41424/q4-2016-total-gpu-shipments.html
110. Revised Data Reveals PC GPU Market Increased 0.2% Quarter-to-Quarter - I-Connect007, accessed May 7, 2025, https://iconnect007.com/article/112617/revised-data-reveals-pc-gpu-market-increased-02-quartertoquarter/112620/ein
111. PC market to grow in double digits - Gartner - Companies | siliconrepublic.com - Ireland's Technology News Service, accessed May 7, 2025, https://www.siliconrepublic.com/business/pc-market-to-grow-in-double-digits-gartner
112. PC Market Up 14% in Q4, Says IDC - Circuits Assembly, accessed May 7, 2025, https://circuitsassembly.com/ca/products-itemid-fix/290-news/2005-news/10075-pc-market-up-14-in-q4-says-idc.html
113. A world of difference - – New Canalys research highlights regional differences in global mobile device market, accessed May 7, 2025, http://www.canalys.com/static/press_release/2004/r2004061.pdf
114. Smart phone shipments break records - The Register, accessed May 7, 2025, https://www.theregister.com/2005/02/01/mobile_devices_q4_04/
115. Smartphones - Beyond Voice to Information and Entertainment | Newswire, accessed May 7, 2025, https://www.newswire.com/news/smartphones-beyond-voice-to-information-and-entertainment-29035
116. World mobile phone shipments up 25% - The Register, accessed May 7, 2025, https://www.theregister.com/2004/10/27/mobile_phone_market_q3_04/
117. Jon Peddie Research reports that overall GPU market was up an astounding 20% – desktop displaced mobile - Graphics - News - HEXUS.net, accessed May 7, 2025, https://m.hexus.net/tech/news/graphics/10245-jon-peddie-research-reports-overall-gpu-market-astounding-20-desktop-displaced-mobile/
118. JPR: Total PC GPU Shipments Increased by 6% From Last Quarter and 20% Year-to-Year, accessed May 7, 2025, https://www.techpowerup.com/319714/jpr-total-pc-gpu-shipments-increased-by-6-from-last-quarter-and-20-year-to-year
119. Global Server Market Grew to $50.9 Billion in 2014 | ServerWatch, accessed May 7, 2025, https://www.serverwatch.com/server-news/global-server-market-grew-to-50-9-billion-in-2014/
120. IDC | Blades Made Simple, accessed May 7, 2025, https://bladesmadesimple.com/tag/idc/
121. Gartner reports PC shipments went up in Q4 2014, but IDC says otherwise, accessed May 7, 2025, https://www.windowscentral.com/gartner-says-pc-shipments-went-q4-2014-idc

-says-they-went-down
122. Gartner contradicts IDC, says worldwide PC shipments grew 1% in Q4 2014, accessed May 7, 2025, https://blog.gsmarena.com/gartner-worldwide-pc-shipments-grew-1-percent-q4-2014-lenovo-leader-pc-shipments/
123. In a Near Tie, Apple Closes the Gap on Samsung in the Fourth Quarter as Worldwide Smartphone Shipments Top 1.3 Billion for 2014, According to IDC | Business Wire, accessed May 7, 2025, https://www.businesswire.com/news/home/20150129005968/en/In-a-Near-Tie-Apple-Closes-the-Gap-on-Samsung-in-the-Fourth-Quarter-as-Worldwide-Smartphone-Shipments-Top-1.3-Billion-for-2014-According-to-IDC
124. Windows Phone Grew Unit Sales in 2014, Lost Market Share - Thurrott.com, accessed May 7, 2025, https://www.thurrott.com/mobile/windows-phone/1588/windows-phone-grew-unit-sales-in-2014-lost-market-share
125. Tablet shipments slip in Q4 for a lackluster 2014 - GeekWire, accessed May 7, 2025, https://www.geekwire.com/2015/tablet-shipments-slip-q4/
126. Tablet sales set to hit the buffers - Headlines, features, photo and videos from ecns.cn|china|news|chinanews, accessed May 7, 2025, http://www.ecns.cn/business/2014/12-29/148498.shtml
127. GPU market up—Intel and Nvidia graphics winners in Q4, AMD down, accessed May 7, 2025, https://www.jonpeddie.com/news/gpu-market-upintel-and-nvidia-graphics-winners-in-q4-amd-down/
128. The State Of PC Graphics Sales Q2 2014 - AnandTech, accessed May 7, 2025, https://www.anandtech.com/show/8446/the-state-of-pc-graphics-sales-q2-2014
129. GPU shipments (MarketWatch) Q2 2014 – Charts and Images - Jon Peddie Research, accessed May 7, 2025, https://www.jonpeddie.com/news/gpu-shipments-marketwatch-q2-2014-charts-and-images/
130. AMD GPU shipments jumped 11 per cent in Q2 2014 say analysts - HEXUS.net, accessed May 7, 2025, https://m.hexus.net/business/news/components/73489-amd-gpu-shipments-jumped-11-per-cent-q2-2014-say-analysts/?print=1
131. AMD vs. Intel CPU Market Share: History and Prediction, accessed May 7, 2025, https://pcviewed.com/amd-vs-intel-cpu-market-share/
132. AMD Dropping Market Share in Desktop Segment | [H]ard|Forum, accessed May 7, 2025, https://hardforum.com/threads/amd-dropping-market-share-in-desktop-segment.2036428/
133. Arm Holdings: Super Growth But Fairly Valued - LongPort, accessed May 7, 2025, https://longportapp.com/en/news/212438722
134. NVIDIA vs. AMD Discrete GPU Market Share (2010 to 2025) - PCViewed, accessed May 7, 2025, https://pcviewed.com/nvidia-vs-amd-discrete-gpu-market-share/

135. Gpu MarketShare ThroughOut The Years- Now, Nvidia 84%, Amd 12%, Intel 4%. - Jon Peddie Research. - Reddit, accessed May 7, 2025, https://www.reddit.com/r/nvidia/comments/14gske5/gpu_marketshare_throughout_the_years_now_nvidia/
136. Worldwide Server Market Revenue Increased 91% During the Fourth Quarter of 2024, according to IDC - Nvidia continues dominating the GPU server space, accessed May 7, 2025, https://www.idc.com/getdoc.jsp?containerId=prUS53264225
137. Arm targets 50 percent of datacenter CPUs this year - The Register, accessed May 7, 2025, https://www.theregister.com/2025/04/01/arm_datacenter_cpu_market/
138. Exclusive: Arm expects its share of data center CPU market sales to rocket to 50% this year, accessed May 7, 2025, https://semiwiki.com/forum/threads/exclusive-arm-expects-its-share-of-data-center-cpu-market-sales-to-rocket-to-50-this-year.22441/
139. Analysts are calling 2023's decline in PC shipments 'unparalleled in the industry's recorded history' but suggest the only way now is up | PC Gamer, accessed May 7, 2025, https://www.pcgamer.com/analysts-are-calling-2023s-decline-in-pc-shipments-unparalleled-in-the-industrys-recorded-history-but-suggest-the-only-way-now-is-up/
140. The PC Market Closed out 2024 with Slight Growth and Mixed Views on What 2025 Will Bring, according to IDC, accessed May 7, 2025, https://www.idc.com/getdoc.jsp?containerId=prUS53061925
141. AMD Sees Growth in Desktop CPU Market Share - The Munich Eye, accessed May 7, 2025, https://themunicheye.com/amd-cpu-market-share-growth-desktop-segment-10761
142. AMD gained consumer desktop and laptop CPU market share in 2024, server passes 25 percent - Reddit, accessed May 7, 2025, https://www.reddit.com/r/hardware/comments/1ioke5m/amd_gained_consumer_desktop_and_laptop_cpu_market/
143. Worldwide Smartphone Forecast Update, 2023–2027: December 2023 - IDC, accessed May 7, 2025, https://my.idc.com/getdoc.jsp?containerId=US51434423
144. Smartphone Market Share - IDC, accessed May 7, 2025, https://www.idc.com/promo/smartphone-market-share/
145. Arm-Based Servers Market to Reach USD 22.79 Billion by 2032, Driven by Cloud Computing, Data Centers, and AI Applications | SNS Insider - GlobeNewswire, accessed May 7, 2025, https://www.globenewswire.com/news-release/2025/05/06/3075220/0/en/Arm-Based-Servers-Market-to-Reach-USD-22-79-Billion-by-2032-Driven-by-Cloud-Computing-Data-Centers-and-AI-Applications-SNS-Insider.html
146. Can Arm's Mobile Lead Translate to AI? Chip Designer Bets on Efficiency | PYMNTS.com, accessed May 7, 2025, https://www.pymnts.com/artificial-intelligence-2/2025/can-arms-mobile-lead-tra

nslate-to-ai-chip-designer-bets-on-efficiency/
147. AMD grabs GPU market share from Nvidia as GPU shipments rise slightly in Q4, accessed May 7, 2025, https://www.tomshardware.com/tech-industry/amd-grabs-a-share-of-the-gpu-market-from-nvidia-as-gpu-shipments-rise-slightly-in-q4
148. Nvidia's GPU sales declined 11% on an annual basis from $3.04B in 2023 to $2.5B in 2024 : r/pcmasterrace - Reddit, accessed May 7, 2025, https://www.reddit.com/r/pcmasterrace/comments/1iz27fb/nvidias_gpu_sales_declined_11_on_an_annual_basis/
149. Latest NVIDIA Statistics in 2025 (Downloadable) | StatsUp - Analyzify, accessed May 7, 2025, https://analyzify.com/statsup/nvidia
150. Data Center GPU Market Size, Share, Industry Report 2032 - MarketsandMarkets, accessed May 7, 2025, https://www.marketsandmarkets.com/Market-Reports/data-center-gpu-market-18997435.html
151. AMD clawed back 7% graphics market share from Nvidia at the end of 2024, but the outlook for the whole industry in 2025 looks iffy | PC Gamer, accessed May 7, 2025, https://www.pcgamer.com/hardware/graphics-cards/amd-clawed-back-7-percent-graphics-market-share-from-nvidia-at-the-end-of-2024-but-the-outlook-for-the-whole-industry-in-2025-looks-iffy/
152. Q4'24 PC GPU shipments increased by 6.2% from last quarter - Jon Peddie Research, accessed May 7, 2025, https://www.jonpeddie.com/news/q424-pc-gpu-shipments-increased-by-4-4-from-last-quarter/
153. accessed December 31, 1969, https://www.gartner.com/en/newsroom/press-releases/2005-01-18-gartner-says-worldwide-pc-shipments-grew-12-percent-in-2004
154. Market Watch – a report series on the Graphics Processor Unit ..., accessed May 7, 2025, https://www.jonpeddie.com/reports/market-watch/
155. accessed December 31, 1969, https://www.palminfocenter.com/news/7613/gartner-worldwide-pda-shipments-grew-7-in-2004/
156. accessed December 31, 1969, https://www.idc.com/getdoc.jsp?containerId=prUS25482415
157. accessed December 31, 1969, https://www.gartner.com/en/newsroom/press-releases/2015-01-12-gartner-says-worldwide-pc-shipments-declined-1-percent-in-fourth-quarter-of-2014
158. accessed December 31, 1969, https://www.idc.com/getdoc.jsp?containerId=prUS25450615
159. accessed December 31, 1969, https://www.jonpeddie.com/press-releases/details/jon-peddie-research-reports-fourth-quarter-gpu-market-shipments
160. accessed December 31, 1969, https://www.idc.com/getdoc.jsp?containerId=prUS51971124

161. accessed December 31, 1969, https://www.gartner.com/en/newsroom/press-releases/2024-01-11-gartner-says-worldwide-pc-shipments-declined-14-8-percent-in-2023-but-grew-in-the-fourth-quarter
162. Apple Grabs the Top Spot in the Smartphone Market in 2023 along ..., accessed May 7, 2025, https://www.idc.com/getdoc.jsp?containerId=prUS51776424
163. accessed December 31, 1969, https://www.jonpeddie.com/press-releases/gpu-shipments-increase-in-q4-2023
164. ROPS, FLOPS, and BOPS - Jon Peddie Research, accessed May 7, 2025, https://www.jonpeddie.com/blog/rops-flops-and-bops/
165. accessed December 31, 1969, https://archive.org/details/byte-magazine-1985-01/page/n5/mode/2up
166. accessed December 31, 1969, https://archive.org/details/creativecomputingv11n01jan1985/page/n1/mode/2up
167. what is the difference/similarity between MIPS and FLOPS? - Super User, accessed May 7, 2025, https://superuser.com/questions/488515/what-is-the-difference-similarity-between-mips-and-flops
168. IBM System/370 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/IBM_System/370
169. MTRioom - NASA Technical Reports Server, accessed May 7, 2025, https://ntrs.nasa.gov/api/citations/19880014800/downloads/19880014800.pdf
170. What Is a Flop? - Nick Higham, accessed May 7, 2025, https://nhigham.com/2023/09/05/what-is-a-flop/
171. Intel 8087 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Intel_8087
172. About: Intel 8087 - DBpedia, accessed May 7, 2025, https://dbpedia.org/page/Intel_8087
173. Minicomputer - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Minicomputer
174. Commodore 64 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Commodore_64
175. IBM Personal Computer - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/IBM_Personal_Computer
176. 1984 - Computer History, accessed May 7, 2025, https://www.computerhope.com/history/1984.htm
177. Influence of the IBM PC on the personal computer market - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Influence_of_the_IBM_PC_on_the_personal_computer_market
178. The IBM PC, accessed May 7, 2025, https://www.ibm.com/history/personal-computer
179. The 'Overpriced' Mac in 1984 - Low End Mac, accessed May 7, 2025, https://lowendmac.com/2005/the-overpriced-mac-in-1984/

180. The Computer History Museum's 40th Anniversary Celebration Of The Macintosh Includes A Shoutout To Brown CS, accessed May 7, 2025, https://blog.cs.brown.edu/2024/02/27/the-computer-history-museums-40th-anniversary-celebration-of-the-macintosh-includes-a-shoutout-to-brown-cs/
181. The IBM System/390, accessed May 7, 2025, https://www.ibm.com/history/system-390
182. IBM 3081, accessed May 7, 2025, https://bitsavers.trailing-edge.com/pdf/datapro/datapro_reports_70s-90s/IBM/70C-491-07_8103_IBM_3081.pdf
183. IBM 308X - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/IBM_308X
184. Apple II Microcomputer | National Museum of American History, accessed May 7, 2025, https://americanhistory.si.edu/collections/object/nmah_334638
185. www.pagetable.com, accessed May 7, 2025, https://www.pagetable.com/?p=547#:~:text=This%20data%2C%20with%20Commodore's%20own,about%20200%2C000%20units%20per%20month.
186. How many Commodore 64 computers were really sold? - pagetable.com, accessed May 7, 2025, https://www.pagetable.com/?p=547
187. First Macintosh Press Release - Stanford University, accessed May 7, 2025, https://web.stanford.edu/dept/SUL/sites/mac/primary/docs/pr1.html
188. Macintosh 128K - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Macintosh_128K
189. Sales numbers for the IBM PC 5150? (or: What was behind the PCjr decision?), accessed May 7, 2025, https://forum.vcfed.org/index.php?threads/sales-numbers-for-the-ibm-pc-5150-or-what-was-behind-the-pcjr-decision.1245051/
190. 1984 | Timeline of Computer History | Computer History Museum, accessed May 7, 2025, https://www.computerhistory.org/timeline/1984/
191. the 3081/e emulator, a processor for use in on - line and off - SciSpace, accessed May 7, 2025, https://scispace.com/pdf/the-3081-e-emulator-a-processor-for-use-in-on-line-and-off-2he2i8imrq.pdf
192. In More Depth: MIPS, MOPS, and Other FLOPS, accessed May 7, 2025, https://course.ccs.neu.edu/cs3650/ssl/TEXT-CD/Content/COD3e/InMoreDepth/IMD4-MIPS-MOPS-and-Other-FLOPS.pdf
193. How much faster is an Intel i7 than an 8 bit 6502? [closed] - Super User, accessed May 7, 2025, https://superuser.com/questions/501458/how-much-faster-is-an-intel-i7-than-an-8-bit-6502
194. 6502.org • View topic - Microcontroller based Math Co-Processor, accessed May 7, 2025, http://forum.6502.org/viewtopic.php?p=102135
195. A Forgotten Consumer PC Becomes A Floating Point Powerhouse | Hackaday, accessed May 7, 2025, https://hackaday.com/2025/02/17/a-forgotten-consumer-pc-becomes-a-floating-point-powerhouse/

196. NEC V20 D70108D VS 8088 CPU Benchmark + Windows 1.01 Booting - YouTube, accessed May 7, 2025, https://www.youtube.com/watch?v=Z1u-IrBT9hE
197. Floating-point arithmetic - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Floating-point_arithmetic
198. Computer History -- Overview at MROB - Robert Munafo, accessed May 7, 2025, http://www.mrob.com/pub/comp/computer-history.html
199. x87 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Intel_80287
200. x87 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/X87
201. What were the most common applications of the 8087? - Retrocomputing Stack Exchange, accessed May 7, 2025, https://retrocomputing.stackexchange.com/questions/16832/what-were-the-most-common-applications-of-the-8087
202. IBM Compatible Computers, accessed May 7, 2025, https://www2.iath.virginia.edu/elab/hfl0108.html
203. CPUs - DOS Days, accessed May 7, 2025, https://www.dosdays.co.uk/topics/cpus.php
204. Motorola 68881 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Motorola_68881
205. Floating Point arithmetic in 68000 - Google Groups, accessed May 7, 2025, https://groups.google.com/g/comp.sys.amiga.programmer/c/6kHVFPXy_QQ
206. FPU Emulator - Page 4 - exxos's Atari, Amiga & retro forum, accessed May 7, 2025, https://www.exxosforum.co.uk/forum/viewtopic.php?t=5619&start=30
207. How does the MicroVax compare to the 68,000 CPU in benchmarks speed? - Quora, accessed May 7, 2025, https://www.quora.com/How-does-the-MicroVax-compare-to-the-68-000-CPU-in-benchmarks-speed
208. 5.11: electrosphere - NYU, accessed May 7, 2025, http://neconomides.stern.nyu.edu/networks/homeworks/Apple_story.htm
209. Intel - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Intel
210. Intel Corporation Annual Report 1994, accessed May 7, 2025, https://www.intel.com/content/dam/doc/report/history-1994-annual-report.pdf
211. AMD: A History of Advanced Marketing Deception - Mr. Robot Computer Repair, accessed May 7, 2025, https://robotatx.com/amd/
212. The History Of AMD CPUs | Tom's Hardware, accessed May 7, 2025, https://www.tomshardware.com/picturestory/713-amd-cpu-history.html
213. PowerPC Architecture Was Not a Failure - Low End Mac, accessed May 7, 2025, https://lowendmac.com/2009/powerpc-architecture-was-not-a-failure/
214. Graphics processing unit - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Graphics_processing_unit
215. GPU Grudge Match '94: ATi vs S3 - Smug and Play, accessed May 7, 2025, https://smugnplay.com/blog/gpu-grudge-match-94-ati-vs-s3
216. S3 Trio64 - Hardware museum, accessed May 7, 2025, https://hw-museum.cz/vga/387/s3-trio64
217. S3 Trio 3D 2X - Hardware museum, accessed May 7, 2025,

http://hw-museum.cz/vga/230/s3-trio-3d-2x
218. ATI Mach64 CT - 512bit.net, accessed May 7, 2025, http://www.512bit.net/ati/ati_mach64_ct.html
219. Comparing GPUs across architectures and tiers | Baseten Blog, accessed May 7, 2025, https://www.baseten.co/blog/comparing-gpus-across-architectures-and-tiers/
220. Nvidia - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Nvidia
221. Motorola Begins Sampling PowerPC 601 - CECS, accessed May 7, 2025, https://www.cecs.uci.edu/~papers/mpr/MPR/ARTICLES/070602.pdf
222. Export Compliance Metrics for Intel® Itanium® and Pentium® Processors, accessed May 7, 2025, https://www.intel.com/content/www/us/en/support/articles/000007250/processors.html
223. Performance Characterization of the Pentium Pro Processor - TAMS, accessed May 7, 2025, https://tams.informatik.uni-hamburg.de/lehre/2001ss/proseminar/mikroprozessoren/papers/pentium-pro-performance.pdf
224. Am486 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Am486
225. The Red Hill CPU Guide: the 486 matures ... and withers away, accessed May 7, 2025, https://www.redhill.net.au/c/c-5.html
226. 41 Year Stock Price History | AMD - Macrotrends, accessed May 7, 2025, https://macrotrends.net/stocks/charts/AMD/amd/stock-price-history
227. Intel Pentium 4 3.00 GHz - CPU Benchmarks, accessed May 7, 2025, https://www.cpubenchmark.net/cpu.php?cpu=Intel+Pentium+4+3.00GHz
228. Intel Pentium 4 HT 3.0E Specs - CPU Database - TechPowerUp, accessed May 7, 2025, https://www.techpowerup.com/cpu-specs/pentium-4-ht-3-0e.c266
229. AMD Athlon 64 3200+ Specs - CPU Database - TechPowerUp, accessed May 7, 2025, https://www.techpowerup.com/cpu-specs/athlon-64-3200.c53
230. Athlon 64 - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Athlon_64
231. accessed December 31, 1969, https://www.tomshardware.com/reviews/cpu-charts-2004,833.html
232. Impact of the New Generation of x86 on the Server Market - Intel, accessed May 7, 2025, https://www.intel.de/content/dam/doc/white-paper/performance-xeon-7500-next-gen-x86-paper.pdf
233. Le Misancalculation: A one act Itanium tragedy by IDC - The Register, accessed May 7, 2025, https://www.theregister.com/2006/02/17/itanic_oracle_idc/?page=2
234. Itanium: A cautionary tale - CNET, accessed May 7, 2025, https://www.cnet.com/tech/tech-industry/itanium-a-cautionary-tale/
235. List of AMD graphics processing units - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/List_of_AMD_graphics_processing_units
236. Architecture of a commodity GPU: NVIDIA GeForce 6800 FX: It has six... - ResearchGate, accessed May 7, 2025,

https://www.researchgate.net/figure/Architecture-of-a-commodity-GPU-NVIDIA-GeForce-6800-FX-It-has-six-programmable-vertex_fig1_221215021
237. GeForce 6 series - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/GeForce_6_series
238. ATI Radeon X800 XT Mac Edition versus GeForce 6800 Ultra and Others - BareFeats, accessed May 7, 2025, https://barefeats.com/radx800.html
239. ATI Radeon X800 XT Mac Edition Graphics Card review - xlr8yourmac.com, accessed May 7, 2025, https://www.xlr8yourmac.com/Graphics/mac_X800_XT_review/index.html
240. NVIDIA GeForce 6800 Ultra Specs | TechPowerUp GPU Database, accessed May 7, 2025, https://www.techpowerup.com/gpu-specs/geforce-6800-ultra.c10
241. NVIDIA GeForce 6800 Ultra Specs | TechPowerUp GPU Database, accessed May 7, 2025, https://www.techpowerup.com/gpu-specs/geforce-6800-ultra.c720
242. ATI Radeon X800 XT Platinum Specs | TechPowerUp GPU Database, accessed May 7, 2025, https://www.techpowerup.com/gpu-specs/radeon-x800-xt-platinum.c23
243. ATi Radeon X800 XT Platinum Edition - Hardware museum, accessed May 7, 2025, https://www.hw-museum.cz/vga/98/ati-radeon-x800-xt-platinum-edition
244. RADEON X800 XT - Price performance comparison - Video Card Benchmarks, accessed May 7, 2025, https://www.videocardbenchmark.net/gpu.php?gpu=RADEON+X800+XT&id=81
245. Radeon X800 series - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Radeon_X800_series
246. accessed December 31, 1969, https://anandtech.com/show/1319
247. accessed December 31, 1969, https://anandtech.com/show/1325
248. Vector Floating-Point (VFP) - ARM11 MPCore Processor Technical Reference Manual r2p0, accessed May 7, 2025, https://developer.arm.com/documentation/ddi0360/latest/introduction/debug-and-programming-support/vector-floating-point--vfp-
249. Advanced Microcontrollers Grzegorz Budzyń Lecture 7: ARM cores, part 3, accessed May 7, 2025, https://www.ue.pwr.edu.pl/advanced_microcontrollers/adv_m_7.pdf
250. Peak FLOPs per cycle for ARM11 and Cortex-A7 cores in Raspberry Pi 1 and 2, accessed May 7, 2025, https://stackoverflow.com/questions/30976071/peak-flops-per-cycle-for-arm11-and-cortex-a7-cores-in-raspberry-pi-1-and-2
251. Intel 'Haswell' Xeon E5s Aimed Squarely at HPC - HPCwire, accessed May 7, 2025, https://www.hpcwire.com/2014/09/08/intel-haswell-xeon-e5s-aimed-squarely-hpc/
252. Detailed Specifications of the Intel Xeon E5-4600 v3 “Haswell-EP” Processors - Microway, accessed May 7, 2025, https://www.microway.com/knowledge-center-articles/detailed-specifications-intel-xeon-e5-4600-v3-haswell-ep-processors/
253. Piledriver (microarchitecture) - Wikipedia, accessed May 7, 2025,

https://en.wikipedia.org/wiki/Piledriver_(microarchitecture)
254. AMD Launches Piledriver-based Opteron 6300 Server Chips - PC Perspective, accessed May 7, 2025, https://pcper.com/2012/11/amd-launches-piledriver-based-opteron-6300-server-chips/
255. Intel Core i7-4770K Review - PCMag, accessed May 7, 2025, https://www.pcmag.com/reviews/intel-core-i7-4770k
256. What is the FLOPS of an Intel Core i7 processor? - Quora, accessed May 7, 2025, https://www.quora.com/What-is-the-FLOPS-of-an-Intel-Core-i7-processor
257. Using a 4.7 GHz FX-8350 in 2023 - AMDs Infamous 8 Core Processor vs Modern Games (RA Tech), accessed May 7, 2025, https://www.reddit.com/r/Amd/comments/13kprpy/using_a_47_ghz_fx8350_in_2023_amds_infamous_8/
258. AMD FX-8350 review - TechRadar, accessed May 7, 2025, https://www.techradar.com/reviews/pc-mac/pc-components/processors/amd-fx-8350-1137704/review
259. Apple A8 - Chaynikam.Info, accessed May 7, 2025, https://www.chaynikam.info/fr/soc62.html
260. CPU Performance - Qualcomm Snapdragon 805 Performance Preview - AnandTech, accessed May 7, 2025, https://www.anandtech.com/Show/Index/8035?cPage=8&all=False&sort=0&page=2&slug=qualcomm-snapdragon-805-performance-preview
261. Qualcomm Snapdragon 805 APQ8084 SoC - NotebookCheck.net Tech, accessed May 7, 2025, https://www.notebookcheck.net/Qualcomm-Snapdragon-805-APQ8084-SoC.108360.0.html
262. NVIDIA GeForce GTX 980 Specs | TechPowerUp GPU Database, accessed May 7, 2025, https://www.techpowerup.com/gpu-specs/geforce-gtx-980.c2621
263. NVIDIA GeForce GTX 980 Review: Does Maxwell Bring Maximum Gameplay? - Techgage, accessed May 7, 2025, https://techgage.com/article/nvidia-geforce-gtx-980-review-does-maxwell-bring-maximum-gameplay/
264. AMD Radeon R9 290X Specs | TechPowerUp GPU Database, accessed May 7, 2025, https://www.techpowerup.com/gpu-specs/radeon-r9-290x.c2460
265. AMD Radeon R9 290 Specs | TechPowerUp GPU Database, accessed May 7, 2025, https://www.techpowerup.com/gpu-specs/radeon-r9-290.c2397
266. A skeptic's take on AI electricity load growth | Latitude Media, accessed May 7, 2025, https://www.latitudemedia.com/news/catalyst-a-skeptics-take-on-ai-electricity-load-growth/
267. The Government's AI plans will supercharge electricity demand | Q4 2024 Quarterly Report, accessed May 7, 2025, https://reports.electricinsights.co.uk/q4-2024/the-governments-ai-plans-will-supercharge-electricity-demand/
268. AI and Semiconductor - a Server GPU Market Analysis and Forecast to 2034,

accessed May 7, 2025, https://www.insightaceanalytic.com/report/ai-and-semiconductor---a-server-gpu-market/2412
269. The Future of the Semiconductor Industry and its People | Blog - Tokyo Electron Ltd., accessed May 7, 2025, https://www.tel.com/blog/all/20250127_001.html
270. 2025 global semiconductor industry outlook - Deloitte, accessed May 7, 2025, https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-telecom-outlooks/semiconductor-industry-outlook.html
271. Epoch AI, accessed May 7, 2025, https://epoch.ai/
272. AI Supercomputers May Run Into Power Constraints by 2030 | PYMNTS.com, accessed May 7, 2025, https://www.pymnts.com/artificial-intelligence-2/2025/ai-supercomputers-may-run-into-power-constraints-by-2030
273. Emerald Rapids - Wikipedia, accessed May 7, 2025, https://en.wikipedia.org/wiki/Emerald_Rapids
274. Intel 'Emerald Rapids' 5th-Gen Xeon Platinum 8592+ Review: 64 Cores, Tripled L3 Cache and Faster Memory Deliver Impressive AI Performance - Tom's Hardware, accessed May 7, 2025, https://www.tomshardware.com/pc-components/cpus/intel-emerald-rapids-5th-gen-xeon-platinum-8592-review-64-cores-320mb-of-l3-and-350w-tdp/3
275. AMD's Genoa CPUs Offer Up to 96 5nm Cores Across 12 Chiplets - HPCwire, accessed May 7, 2025, https://www.hpcwire.com/2022/11/10/amds-4th-gen-epyc-genoa-96-5nm-cores-across-12-compute-chiplets/
276. AMD Genoa and Intel Sapphire Rapids review - chpc.utah.edu, accessed May 7, 2025, https://www.chpc.utah.edu/documentation/white_papers/cpus_may2023_v3.pdf
277. Apple M3 Max 40-Core GPU - Notebookcheck, accessed May 7, 2025, https://www.notebookcheck.net/M1-8-Core-GPU-vs-M3-Max-30-Core-GPU-vs-M3-Max-40-Core-GPU_10552_12034_11636.247598.0.html
278. Apple reveals M3 Ultra, taking Apple silicon to a new extreme, accessed May 7, 2025, https://www.apple.com/newsroom/2025/03/apple-reveals-m3-ultra-taking-apple-silicon-to-a-new-extreme/
279. Snapdragon X Elite | Our Newest Laptop Processor - Qualcomm, accessed May 7, 2025, https://www.qualcomm.com/products/mobile/snapdragon/laptops-and-tablets/snapdragon-x-elite
280. Nuvia-based Snapdragon X Elite GPU benchmarks appear in database | Tom's Hardware, accessed May 7, 2025, https://www.tomshardware.com/pc-components/gpus/nuvia-based-snapdragon-x-elite-gpu-benchmarks-appear-in-database
281. State of Windows on Arm, Year End 2024 - Signal65, accessed May 7, 2025, https://signal65.com/research/state-of-windows-on-arm-year-end-2024/

282. The Rise of RISC-V: Is It a Threat to ARM and x86? (Market Growth Stats) | PatentPC, accessed May 7, 2025, https://patentpc.com/blog/the-rise-of-risc-v-is-it-a-threat-to-arm-and-x86-market-growth-stats
283. System On Module Market Growth And Industry Analysis Report 2025-2034, accessed May 7, 2025, https://www.thebusinessresearchcompany.com/report/system-on-module-global-market-report
284. RISC-V Tech Market Size, Share & Growth Report 2032 - SNS Insider, accessed May 7, 2025, https://www.snsinsider.com/reports/risc-v-tech-market-6884
285. [News] China's Push for Chip Independence: Can RISC-V Challenge x86 and Arm's Dominance? - TrendForce, accessed May 7, 2025, https://www.trendforce.com/news/2025/03/14/news-chinas-push-for-chip-independence-can-risc-v-challenge-x86-and-arms-dominance/
286. GPU Benchmarks NVIDIA RTX 4090 vs. NVIDIA RTX 6000 Ada - Bizon Tech, accessed May 7, 2025, https://bizon-tech.com/gpu-benchmarks/NVIDIA-RTX-4090-vs-NVIDIA-RTX-6000-Ada/637vs640
287. GPU Benchmarks NVIDIA A100 80 GB (PCIe) vs. NVIDIA RTX 4090 - Bizon Tech, accessed May 7, 2025, https://bizon-tech.com/gpu-benchmarks/NVIDIA-A100-80-GB-(PCIe)-vs-NVIDIA-RTX-4090/624vs637
288. AMD Radeon 7900 XT/XTX Inference Performance Comparisons : r/LocalLLaMA - Reddit, accessed May 7, 2025, https://www.reddit.com/r/LocalLLaMA/comments/191srof/amd_radeon_7900_xtxtx_inference_performance/
289. AMD Radeon RX 7900 XTX review (Page 4) - www.guru3d.com, accessed May 7, 2025, https://www.guru3d.com/review/amd-radeon-rx-7900-xtx-review/page-4/
290. Cloud computing and AI: 5 predictions for investors - Julius Baer, accessed May 7, 2025, https://www.juliusbaer.com/en/insights/future-insights/digital-disruption/cloud-computing-and-ai-5-predictions-for-investors/
291. Next Generation Computing Market Size, Forecasts 2025-2034, accessed May 7, 2025, https://www.gminsights.com/industry-analysis/next-generation-computing-market
292. Since 2010, the training computation of notable AI systems has doubled every six months, accessed May 7, 2025, https://ourworldindata.org/data-insights/since-2010-the-training-computation-of-notable-ai-systems-has-doubled-every-six-months
293. How much of AI progress is from scaling compute? And how far will it scale? - Metaculus, accessed May 7, 2025, https://www.metaculus.com/notebooks/10688/
294. Graphic Processing Unit Market Size, Share and Forecast 2034 - Zion Market Research, accessed May 7, 2025,

https://www.zionmarketresearch.com/report/graphic-processing-unit-market
295. Top 3 Predictions for 2025: The Future of AI, Cybersecurity, and Digital Identity - Cerby, accessed May 7, 2025, https://www.cerby.com/resources/blog/top-3-predictions-for-2025
296. Global High Performance Computing (HPC) and AI Accelerators Report: Market Size and Growth Projections (2025-2035), Investment Outlook and Opportunities - GlobeNewswire, accessed May 7, 2025, https://www.globenewswire.com/news-release/2025/05/06/3074771/28124/en/Global-High-Performance-Computing-HPC-and-AI-Accelerators-Report-Market-Size-and-Growth-Projections-2025-2035-Investment-Outlook-and-Opportunities.html
297. IDC Estimates Global Spending on Edge Computing to Grow at 13.8% Reaching Nearly $380 Billion by 2028, accessed May 7, 2025, https://www.idc.com/getdoc.jsp?containerId=prUS53261225