Global FLOPs Sold: 1984–2044

Introduction

FLOPs Sold refers to the total **floating-point operations per second (FLOPs)** capacity of all computing hardware **shipped in a given year**. In other words, it is the sum of the peak computational power (in FLOPs) of every computer, server, microprocessor, and accelerator device sold worldwide during that year. Measuring this is challenging – data must be pieced together from historical sales, technical specs, and estimates . Nevertheless, researchers have attempted to inventory the world's computing capacity. For example, Hilbert & López (2012) estimated that by 2007 the world's installed computers could perform on the order of 2×10^20 instructions per second (which is comparable to FLOPs), and further analysis suggests global hardware in 2015 could perform up to ~1.5×10^21 FLOPs . Using such studies, along with industry reports, we can **estimate the FLOPs added each year** and how they break down by architecture.

Below, we analyze key years (1984, 1994, 2004, 2014, 2024) and give projections for 2034 and 2044. For each year, we break down the contribution by major computing architectures: in early years, **mainframes**, **minicomputers**, and **PCs** (with PCs split into Intel-based vs. other CPUs); in later years, **Intel-compatible (x86) CPUs** (including AMD), **ARM-based chips**, and **GPUs** (graphics processors, split into NVIDIA vs. others). We provide tables to compare FLOPs across years and architectures. (All values are rough estimates – where hard data was unavailable, we indicate the reasoning behind the estimate.)

1984: Mainframes, Minis, and the PC Revolution

1984 was a turning point when personal computers (PCs) began to rival traditional mainframes in market significance. In fact, 1984 was the first year that **desktop computer** sales (about $11.6 billion) slightly exceeded **mainframe** sales (around $11.4 billion) . This economic shift implies that the computing capacity sold in PCs was becoming comparable to that in mainframes, given that PCs were much cheaper per FLOP.

- **Mainframes:** These were large enterprise systems (e.g. IBM System/370 series) costing hundreds of thousands of dollars each. A mid-1980s mainframe might deliver on the order of **5–10 million instructions per second** (MIPS), roughly a few **million FLOPs** per second (MFLOPs) of peak performance per system (since many scientific calculations relied on specialized hardware or vector units). Mainframe shipments in 1984 were relatively small in number (hundreds worldwide), but each carried significant performance. We estimate mainframes accounted for on the order of **10^7–10^8 FLOPs** (tens of millions of FLOPs) of the year's shipped capacity.

- **Minicomputers:** Minicomputers (smaller multi-user systems like DEC VAX miniseries) were also important. A typical minicomputer in 1984 (e.g. a VAX-11/780 or newer models) offered around **0.5–2 MIPS** (comparable MFLOPs in peak) and was sold to businesses and labs in the thousands of units. Collectively, minicomputers likely contributed a few times 10^7 FLOPs of new capacity in 1984 (comparable to mainframes in total, as minis were displacing the lower-end mainframe market).
- **Personal Computers (PCs):** The early 1980s saw rapid growth in PC sales. By 1984, the IBM PC and clones ("Intel architecture" PCs using 16-bit **x86 CPUs**) were selling about **2 million units per year**, roughly matching the volume of popular 8-bit home computers like the Commodore 64 . In total, including Apple computers and others, on the order of **5–6 million personal computers** were sold in 1984. However, each PC's performance was modest. An Intel 8088 CPU (IBM PC) ran at 4.77 MHz (~0.33 MIPS), and even the newer 80286 (IBM AT, introduced 1984) at 6–8 MHz achieved around 1–2 MIPS. Many PCs lacked a math co-processor, so floating-point math was slow (effectively a fraction of a MFLOP in software). Nonetheless, with millions of PCs, their **aggregate** added computing power was significant. We estimate **PCs (Intel x86)** contributed roughly **5×10^6 MIPS** (≈5×10^12 operations/sec) of capacity – on the order of **10^7–10^8 FLOPs** – comparable to or exceeding mainframe+mini combined that year.
    - **Intel vs. Other PC CPUs:** The majority of 1984 PCs used Intel's 8088/8086/80286 or compatible chips (some were second-sourced by AMD or others under license, but architecturally the same). **Intel architecture PCs** likely contributed ~70–80% of PC FLOPs that year. **Other chips** made up the rest: for instance, the Apple Macintosh (introduced 1984) and many "home computers" used **Motorola 68000, MOS 6502, Zilog Z80**, etc. These non-Intel PCs were perhaps 1–2 million units (e.g. Commodore, Apple, Atari) but often with lower clocks, so maybe ~20–30% of PC FLOPs. In sum, Intel-based PCs delivered on the order of **10^7 FLOPs**, and non-Intel microcomputers a few ×10^6 FLOPs in 1984.

Summary 1984: Around **10^8 FLOPs total** computing capacity was sold worldwide in 1984 (tens of millions of FLOPs). Mainframes and minis made up roughly one-third of that total, while the burgeoning PC sector contributed about two-thirds (with Intel/x86 PCs dominating that segment). This marked the rising importance of microprocessor-based computing. *(Basis:* Historical sales and typical MIPS/FLOPs per machine; by 1984 PCs outsold mainframes in both units and total compute delivered *.)*

1994: PCs Dominate, RISC and GPUs Emerge

By **1994**, personal computing and microprocessors had decisively overtaken legacy systems in annual computing capacity added. The **total FLOPs sold in 1994** is estimated on the order of **10^15 FLOPs** (a few quadrillion FLOPs) of peak capacity – about 10,000× greater than a decade earlier, thanks to *both* faster chips and higher unit volumes. Several architectural categories contributed:

- **x86 CPUs (Intel-compatible):** The **x86 architecture** (Intel and compatible processors) was the cornerstone of computing by 1994. Roughly **40–50 million PCs** were sold worldwide that year (the PC market was nearing 50M units/year by 1995). Typical 1994 PCs used **Intel 80486** or early **Pentium** processors. For example, an Intel 80486 DX2 ran at 66 MHz ~ delivering ~50 MIPS (around 5–10 MFLOPs with its built-in FPU), while the new Pentium (60–90 MHz) could exceed 100 MIPS (~20 MFLOPs). With tens of millions of such chips, **x86 CPUs** contributed the bulk of 1994's computing. We estimate on the order of **10^15 FLOPs** from x86 shipments alone. Intel had the lion's share, but **AMD** (and Cyrix) also shipped x86-compatible CPUs (AMD's 486 and early Am5x86 chips) – these *Intel-compatible* chips are included in the x86 total. In 1994, x86 likely constituted **the majority (≈80–85%) of new FLOPs** sold globally, reflecting the dominance of PCs and x86 servers.
- **Other CPU Architectures:** A significant minority of computing in 1994 came from non-x86 processors. **RISC architectures** were on the rise in workstations and high-end servers: e.g. **Sun SPARC, DEC Alpha, IBM POWER, HP PA-RISC**, and others powered Unix workstations/servers (though in unit terms, only tens of thousands sold, each delivering hundreds of MFLOPs). **Apple's Macintosh** had transitioned to **Motorola/IBM PowerPC** RISC chips by 1994, contributing a few percent of the PC market. Meanwhile, **mainframes** had declined drastically in relative contribution – IBM's mainframe sales were a smaller niche by the '90s (IBM's revenue crisis in the early '90s underscored this). Still, a high-end 1994 mainframe (IBM System/390) could deliver **hundreds of MIPS**, and IBM sold a few hundred such systems, so mainframes added perhaps **10^12–10^13 FLOPs** (a tiny fraction of the year's total). In sum, **non-x86 CPUs** (RISC workstations, mainframes, older minis, PowerPC Macs, etc.) might account for ~15–20% of 1994 FLOPs sold. This category provided specialized or high-end computing but was being eclipsed by the economics of x86.
- **ARM and Embedded: ARM** architecture was in its infancy on the global stage. ARM chips were used in a few niche areas by 1994 – e.g. early PDAs (the Apple Newton used an ARM 610 in 1993) – but volumes and performance were low. Most consumer electronics still used 8-bit or 16-bit microcontrollers (8051, Z80, etc.) or simple DSPs. ARM's contribution to total FLOPs in 1994 was **negligible (<<1%)**.
- **Graphics Processors (GPUs):** The mid-1990s also saw the appearance of the first **GPU**-like chips, though not yet a major factor in general computing capacity. 1994 was just before the 3D graphics boom – chips like the S3 ViRGE (1995) and 3dfx Voodoo (1996) were imminent. In 1994, PCs did have graphics controllers (VGA cards, etc.), but these were mostly 2D accelerators with no significant floating-point compute capability (no

programmable shaders yet). Thus, **GPUs** contributed ~**0% of FLOPs** in 1994. (Any graphics chips then were fixed-function and not counted as general FLOPs.)

Summary 1994: Total new computing sold in 1994 is on the order of **5×10^15 FLOPs** (several PFLOPs). **Intel-compatible (x86) CPUs** utterly dominated, providing roughly 80%+ of that capacity, as PC shipments exploded. **RISC/Unix and other non-x86 CPUs** provided most of the remainder (high-perf workstations and legacy systems), while **ARM was minor** and **GPUs negligible** in terms of FLOPs. This year exemplified the **PC era** – mass-produced microprocessors driving most computing growth.

2004: Rise of Mobile and Graphics Compute

By **2004**, global computing capacity shipments had grown by another ~300× over 1994, reaching roughly **10^18 FLOPs** (an **exaflop** scale of peak performance added that year). The landscape had further diversified: traditional PC CPUs were still central, but graphics processors and mobile/embedded chips (many ARM-based) were becoming significant.

- **x86 (Intel & AMD):** The x86 architecture remained the single biggest contributor in 2004. Around **180 million PCs** were sold in 2004 (desktop and laptop combined), plus millions of x86-based servers. Intel's Pentium 4 and Xeon CPUs (3–3.6 GHz, with SIMD extensions) could deliver on the order of **~6–10 GFLOPs each** (using SSE2 for floating point). AMD's 64-bit Opteron and Athlon64 chips had similar GFLOP ratings and were gaining market share in both PCs and servers by 2004. If we assume ~200 million x86 CPUs shipped (including multi-core counted as one package) at ~5 GFLOPs average, that's on the order of **10^18 FLOPs** right there. In reality, not every CPU was running at peak or had a vector unit engaged, but peak capacity of x86 shipped this year can be estimated in the low **exaflop range**. Thus, **x86 CPUs** likely contributed on the order of **10^18 FLOPs** (perhaps ~60–70% of total new FLOPs in 2004). Within x86, Intel still held majority of units, but **AMD's portion** (especially with its 64-bit chips) was non-trivial – on the order of 15–20% of x86 FLOPs.
- **ARM and Other Mobile/Embedded:** The **ARM architecture** had quietly become pervasive in mobile and embedded devices by 2004, although the performance per chip was much lower than that of PC CPUs. Nokia, for instance, was shipping hundreds of millions of mobile phones per year in the early 2000s – many of these phones contained ARM7/ARM9 family microcontrollers running at tens of MHz. A single 2004-era ARM9 CPU (e.g. in a feature phone) might deliver **~0.1 MIPS/MHz**, so at 100 MHz that's ~10 MIPS (and perhaps a few MFLOPs). However, the **volume** of ARM-based chips was huge: not only phones, but also iPods, cameras, and myriad embedded systems. By 2004, roughly **1–2 billion ARM-based chips** were being shipped annually (a number that would only grow). Collectively, ARM chips likely contributed on the order of **10^16–**

10^17 FLOPs in 2004 (a few percent of the total) – relatively small in share of FLOPs due to low per-chip performance, but a harbinger of the mobile era. Other embedded CPUs (non-ARM microcontrollers, DSPs) added further, but individually their FLOPs were tiny; in aggregate they form part of this "embedded" contribution. (Notably, by 2007 the **world's application-specific chips** like microcontrollers accounted for *32× the compute capacity of general-purpose PCs* , but in 2004 this shift was underway, not yet complete.)

- **GPUs – NVIDIA vs. Others: Graphics Processing Units** had become a major factor by 2004, especially for consumer graphics and gaming. GPUs are highly parallel and excel at FLOPs, often an order of magnitude more than contemporary CPUs. In 2004, **NVIDIA** and **ATI (now AMD)** were the two main GPU vendors. A high-end GPU like the *NVIDIA GeForce 6800 Ultra* (2004) could perform roughly **40 GFLOPs** (single-precision) and ATI's Radeon X800 similarly offered tens of GFLOPs. That year, tens of millions of discrete GPUs were sold (gaming PCs, workstations) and even more GPUs were integrated in consoles (e.g. Sony PS2's Graphics Synthesizer) or motherboard chipsets. Although these GPUs were primarily used for graphics, their theoretical FLOPs count toward capacity. We estimate **GPUs in 2004 contributed on the order of 10^17 FLOPs** in total (perhaps 10% of global FLOPs sold). NVIDIA had a slight market lead in discrete PC GPUs (roughly 60% share) so **NVIDIA GPUs** might account for ~6×10^16 FLOPs vs. ~4×10^16 FLOPs from **"other" GPUs** (primarily ATI, plus a small number of others like Matrox, etc.). This was also the era GPUs began to be used experimentally for general-purpose computing (GPGPU), though that trend would boom later.
- **Other CPUs (Legacy and RISC):** Outside x86 and ARM, a sliver of 2004's compute came from other architectures. **IBM's PowerPC and Power** processors were used in IBM's servers and some supercomputers, as well as in the Apple Macintosh line (which in 2004 still used PowerPC G4/G5 CPUs). **Sun's SPARC** and **HP/Intel Itanium** were also present in enterprise servers. These RISC/EPIC architectures had small unit share but high per-chip performance (each in the tens of GFLOPs). Additionally, **IBM mainframes** (zSeries) continued to ship in small volumes – each modern mainframe CPU chip by 2004 could perform several GFLOPs, but their total contribution was minimal. Altogether, these "other" architectures likely made up only a few percent of 2004's FLOPs (perhaps ~10^16 FLOPs total).

Summary 2004: Total new computing capacity sold ~**6×10^17–1×10^18 FLOPs** (roughly an **exaflop** of peak power). **x86 CPUs** still delivered the majority (~60%+). **GPUs** emerged with ~10% (NVIDIA being the largest single GPU contributor). **ARM-based chips** and other embedded processors provided a growing slice (~5–10%), while **other legacy/RISC CPUs** filled the remaining few percent. The 2004 landscape shows the **bifurcation** of computing: power was no longer just in PCs/servers, but also in specialized processors like GPUs and the billions of low-cost embedded CPUs.

2014: The Mobile & Accelerator Era

By **2014**, a decade later, the annual FLOPs shipped globally had exploded by another factor of ~1000, reaching on the order of **10^20 FLOPs** (hundreds of exaFLOPs) of capacity. Two major shifts marked this period: the **mobile revolution** (billions of smartphones) powered by ARM chips, and the rise of **accelerators (GPUs)** for high-performance and parallel computing (especially driven by graphics and the dawn of AI workloads). Meanwhile, x86 remained a workhorse for PCs and data centers, but its relative share of total FLOPs began to diminish.

- **x86 (Intel & AMD):** In 2014, x86 CPUs were at their performance peak per core, but unit growth was slowing. About **315 million PCs** were shipped in 2014 (this was near the peak PC era, before declines in subsequent years), plus millions of x86 server processors. Intel's latest CPUs (Core i5/i7 for PCs, Xeon for servers) ran in the few-GHz range but added more cores and wider vectors. A typical 2014 high-end desktop CPU (e.g. Intel Haswell quad-core) could achieve ~100 GFLOPs (using AVX2 256-bit vector units). Server CPUs with more cores could top that. AMD's x86 presence had shrunk on the PC side (though it still supplied CPUs for gaming consoles in 2014, e.g. the PlayStation 4 and Xbox One used 8-core AMD x86 APUs). We estimate the **aggregate x86 FLOPs** shipped in 2014 at roughly **10^19–10^20 FLOPs** (tens of exaFLOPs). This includes millions of PC CPUs (~10^8 units * tens of GFLOPs each) and server CPUs (which, though fewer in number, had higher per-chip FLOPs). **Intel** remained dominant, supplying ~85–90% of these x86 FLOPs, with **AMD** covering most of the remainder.
- **ARM (Mobile/Embedded):** By 2014, **ARM-based chips** were ubiquitous – not only in virtually every smartphone and tablet, but also in embedded IoT devices, consumer electronics, and even some servers. That year, over **1.2 billion smartphones** were sold, nearly all running multi-core ARM SoCs (System-on-Chip). A 2014 smartphone SoC (e.g. Apple A8 or Qualcomm Snapdragon 805) had a **4× ARM Cortex-A** CPU cluster delivering around **20–50 GFLOPs**, plus a GPU (and DSPs) on chip. Moreover, countless ARM microcontrollers (Cortex-M series) were shipping for IoT. All told, ARM partners were shipping on the order of **10+ billion chips per year** by the mid-2010s. In fact, by 2020 ARM was shipping **more chips per quarter (6.7 billion in Q4 2020) than x86 did in a year**, outselling "all other CPU architectures combined" in unit terms . In 2014, ARM's share of **FLOPs** was still behind x86 (because each chip had lower performance), but the sheer volume made it significant. We estimate ARM-based devices contributed on the order of **10^19 FLOPs** in 2014 – roughly **15–20% of total FLOPs**. This includes the cumulative power of billions of mobile SoCs and embedded processors. Notably, **Apple's A-series** ARM chips in iPhones/iPads were performance leaders in mobile, and companies like Qualcomm, Samsung, and MediaTek shipped masses of ARM cores worldwide. The era of mobile computing meant ARM had become a **first-class computing platform** alongside x86.
- **GPUs (NVIDIA and others):** 2014 also fell in the midst of a **GPU computing boom**. GPUs were not only critical for graphics (PC gaming, console hardware, etc.), but were increasingly used for general-purpose computing and artificial intelligence. NVIDIA's CUDA ecosystem (launched 2007) had by 2014 made GPUs a mainstay in scientific and high-performance computing. For example, the world's fastest supercomputer in 2014 (China's Tianhe-2) used Intel's MIC accelerators, but the trend was shifting to GPU-

accelerated systems – many Top500 supercomputers that year leveraged NVIDIA Tesla GPU accelerators, and GPUs contributed a **majority of the FLOPs growth** on the Top500 list around that time . In consumer tech, 2014's GPUs were extremely powerful: an NVIDIA **GeForce GTX 980** (2014) packed ~5 TFLOPs (single precision), and AMD's Radeon R9 290X delivered a similar ~5 TFLOPs. Millions of these discrete GPUs were sold for gaming and workstations. In addition, game consoles (the PS4 and Xbox One, launched 2013) each contained an AMD GPU of ~1.8 TFLOPs, contributing further to non-NVIDIA GPU totals. **NVIDIA** had about 70–80% share of the discrete GPU market in 2014, while **AMD (ATI)** held most of the rest. There were also emerging specialized processors for mining (Bitcoin ASICs) in 2014, but those are application-specific and not counted here. Summing up, the **total GPU FLOPs** shipped in 2014 likely reached **10^19 FLOPs or more**, rivaling the CPU contribution. In percentage, GPUs might represent ~30–40% of 2014's new FLOPs. **NVIDIA's share** of that would be on the order of 20–30% (as its high-end cards and HPC accelerators dominated many segments), with **non-NVIDIA GPUs** (AMD Radeon in PCs/consoles, plus mobile GPUs like Qualcomm Adreno, Imagination PowerVR in phones) making up ~10–15%. *(Mobile GPUs are integrated on ARM SoCs, so one could count their FLOPs under ARM as well; here we include them implicitly in ARM's total or as part of "other GPUs" as appropriate.)*

- **Other Architecture:** By 2014, most formerly significant CPU architectures outside x86/ARM had faded or specialized. IBM's **Power** architecture still powered IBM's UNIX servers and the No. 2 supercomputer (Sequoia) in 2014, and Intel's **Itanium** lingered in a niche. These contributions were comparatively small (perhaps a few ×10^17 FLOPs total, <1% share). One noteworthy new category in the late 2000s/2010s was custom **ASIC accelerators**. For example, by 2011 Google had started using custom application-specific integrated circuits (though its famous **TPU** for AI would only arrive in 2015). In 2014, however, such accelerators were not yet a major portion of global FLOPs; most AI training still happened on GPUs or x86.

Summary 2014: The world shipped on the order of **1×10^20 FLOPs** of computing capacity (≈100 exaFLOPs). **x86 CPUs** were about one-third of that total (still huge, powering PCs and servers). **ARM-based chips** (smartphones, tablets, embedded) contributed roughly another 20%. **GPUs** (led by NVIDIA) accounted for a solid ~30% or more – reflecting how accelerators had become essential for performance. The remainder (~10–15%) came from other or hybrid categories. In short, by 2014 computing was **highly multi-architecture**: the majority of FLOPs sold were outside of traditional PC CPUs, with mobile and GPUs taking the lead in growth.

2024: Present Day – Heterogeneous Computing Explosion

As of **2024**, the annual global hardware shipments likely provide **on the order of 10^22 FLOPs** (tens of *zetta*FLOPs) of peak capacity. The last decade's trends intensified: mobile and IoT devices further increased the share of ARM/RISC processors, while the rise of deep learning created insatiable demand for GPUs and new accelerators. x86 remains important but no longer the majority by FLOPs. We break down 2024 roughly as follows:

- **x86 (Intel & AMD):** The x86 architecture in 2024 is used primarily in **PCs, servers, and cloud data centers**. PC shipments are around 250 million units/year (slightly down from a decade ago), but server CPU shipments grew as cloud computing expanded. Modern x86 chips (e.g. Intel 12th/13th Gen "Core" CPUs, AMD Ryzen and EPYC CPUs) feature many cores and wide vector engines. A single high-end server CPU in 2024 (AMD EPYC or Intel Xeon) can deliver **~1 TFLOP** or more of double-precision, and several TFLOPs in lower precision. Despite this progress, x86's share of total FLOPs has fallen in percentage terms. We estimate **x86 CPUs contribute on the order of 10^21 FLOPs** in 2024, roughly **10–15%** of the world's new compute capacity. Intel's share of this is still substantial, though Intel's data center dominance has been challenged by AMD (which in 2024 has ~20–30% of server CPUs) and by customers adopting ARM-based server chips. Notably, **x86 lost its monopoly in personal computing**: Apple's move to ARM for Macs (starting 2020) and the rise of ARM-based Windows laptops (Snapdragon) have slightly reduced x86 volumes, though x86 still powers the majority of PCs and nearly all traditional laptops/desktops.
- **ARM (and RISC-V) CPUs:** The **ARM architecture** now utterly dominates **consumer electronics and IoT**. Annual unit shipments of ARM-based chips exceeded **25 billion** by the late 2010s , and have continued to grow. In smartphones alone, around **1.4 billion units/year** (2024 figure) each contain a multi-core ARM SoC capable of tens to hundreds of GFLOPs (thanks to advanced 5 nm silicon and built-in AI accelerators). ARM is also making inroads into laptops (Apple's M1/M2 chips deliver **>5 TFLOPs** each using integrated GPUs) and servers (e.g. Amazon's Graviton and Ampere Altra CPUs). Moreover, **embedded and IoT** devices number in the tens of billions per year, including both ARM Cortex-M microcontrollers and emerging **RISC-V** microcontrollers. (By 2024, RISC-V, an open ISA, has begun to capture a portion of embedded markets; it's estimated that billions of RISC-V cores ship in devices like appliances, wearables, etc., though typically with modest performance – we include them in this category as "ARM/RISC-V" aggregate). Given the massive quantity, the **aggregate FLOPs of ARM/RISC chips** in 2024 is enormous – likely on par with or exceeding x86. We estimate ARM/RISC-based CPUs contribute around **10^21–10^22 FLOPs**, roughly **20–30% of the total**. For example, the sum of all smartphone chip FLOPs (each smartphone often has ~0.5–1 TFLOP including CPU, GPU, NPU) plus all microcontrollers (tiny each, but billions of them) plus ARM servers, etc., yields a very large number. It's safe to say that *in 2024, non-x86 CPUs (mostly ARM) account for the largest share of general-purpose processors by volume and a very large share by FLOPs*. This is a dramatic change from a few decades ago – a reflection of the **"computing everywhere"** reality (in everything from thermostats to cars).
- **GPUs (NVIDIA and others):** If one single category can claim the **majority of FLOPs shipped in 2024**, it is **GPUs and accelerators**. The explosion of AI (especially deep learning) in the late 2010s and 2020s led to an arms race in GPU capability. NVIDIA's GPU lineup in 2024 (e.g. the **NVIDIA A100/A1000** series and newer H100) deliver **hundreds of TFLOPs** *each* (when using tensor cores or lower precision). These chips are deployed by the thousands in AI training clusters and supercomputers. NVIDIA has

reported record data center revenue, reflecting massive deployment of AI GPUs – indeed, in the cloud & data center AI processor market, **NVIDIA holds about 80% share as of 2020** and likely similar or greater by 2024. In addition to data center GPUs, millions of consumer GPUs (GeForce and Radeon cards of 2024) provide tens of TFLOPs each for gaming and graphics. The **console market** remains strong (Sony PS5 and Microsoft Xbox Series X, launched 2020, each have ~10 TFLOPs AMD GPUs, with 30+ million units of each sold by 2024 – boosting the "non-NVIDIA GPU" count). There are also AI-specific chips now: **Google's TPU** (not a GPU but a matrix accelerator) has been produced in multiple generations for Google's internal use; other companies (Graphcore, Habana/Intel, etc.) also ship AI accelerators. We include these in the broad "GPU/accelerator" category. Taking all this into account, the total FLOP/s capacity of **accelerators shipped in 2024** likely reaches into the **tens of zettaFLOPs**. For instance, just a few thousand NVIDIA H100 GPUs (<0.001% of units) already provide an exaflop-scale cluster. Globally, millions of GPUs across all vendors are shipped each quarter. We estimate **GPUs/accelerators are ~50% of 2024's FLOPs**. In particular, **NVIDIA's GPUs** alone might be ~40% of total FLOPs (given their dominance in high-performance sectors), with **non-NVIDIA GPUs** (including AMD's Radeon in PCs and consoles, plus AI chips from others) making up ~10%. While these percentages are rough, it's clear that **the balance of compute has shifted strongly toward parallel accelerators** – a modern GPU chip can have 10x-100x the FLOPs of a CPU, and spending on GPUs now rivals or exceeds spending on CPUs in many contexts. *(For context, one analysis estimated total global computing in 2015 was ~2×10^20–1.5×10^21 FLOPs ; by 2023, just GPUs+TPUs were estimated at ~3.98×10^21 FLOPs in operation . In 2024 alone, new GPU shipments added many times that.)*

- **Other/Quantum:** Other architectures in 2024 are a tiny fraction. IBM still ships a few **mainframes** (z16, etc.) which have impressive per-chip performance (billions of FLOPs) but negligible in the global sum. **IBM Power** and **Oracle SPARC** exist for legacy enterprise systems but in very low volumes. One emerging area is **quantum computers**, but these are not measured in FLOPs and remain experimental with only dozens of stable qubits – not a contributor to classical FLOPs. Overall, beyond CPUs and GPUs, no other architecture exceeds a few percent of the global FLOPs share in 2024.

Summary 2024: On the order of **5×10^22 FLOPs** (tens of zettaflops) of computing power were sold in 2024. The **single largest component** is now **GPUs/AI accelerators** (~50% of the total, with NVIDIA by far the biggest player). **ARM and similar CPUs** (for mobile/IoT) also account for a very large chunk (~25% or more), reflecting billions of devices per year . **x86 CPUs**, while still critical for many uses, likely provide under 20% of the raw FLOPs added. This represents a striking inversion from the 1990s: the majority of computing capacity now lies outside traditional PC/server processors. Table 1 summarizes the historical data from 1984–2024 across architectures.

Table 1. Estimated FLOPs Sold Globally by Year and Architecture (peak hardware capacity shipped during that year, in scientific notation). Each entry is the approximate share of that year's total FLOPs attributable to a given architecture category. (*PC* refers to personal computers; *GPU* refers to graphics processors; **Note:** "Other CPUs" includes legacy architectures like mainframes, minis, RISC/Unix systems, etc., that are not x86 or ARM.)

Category	1984	1994	2004	2014	2024
Mainframes & Minis	~3×10^7 FLOPs (≈30%)	~1×10^13 FLOPs (≲5%)	~5×10^15 FLOPs (≲1%)	~1×10^17 FLOPs (<1%)	~1×10^18 FLOPs (<1%)
x86 CPUs (Intel/AMD)	~6×10^7 FLOPs (≈60%)	~4×10^15 FLOPs (≈85%)	~4×10^17 FLOPs (≈70%)	~3×10^19 FLOPs (≈30%)	~8×10^21 FLOPs (≈15%)
Other CPUs (Non-x86)	~1×10^7 FLOPs (≈10%)	~6×10^14 FLOPs (≈10%)	~2×10^16 FLOPs (≈3%)	~5×10^17 FLOPs (≈5%)	~5×10^20 FLOPs (≈1%)
ARM CPUs	~0 (0%)	~0 (0%)	~5×10^16 FLOPs (≈5%)	~2×10^19 FLOPs (≈20%)	~1.2×10^22 FLOPs (≈25%)
GPUs – NVIDIA	~0 (0%)	~0 (0%)	~6×10^16 FLOPs (≈10%)	~3×10^19 FLOPs (≈30%)	~2×10^22 FLOPs (≈40%)
GPUs – Other	~0 (0%)	~0 (0%)	~4×10^16 FLOPs (≈7%)	~1×10^19 FLOPs (≈10%)	~5×10^21 FLOPs (≈10%)
Total FLOPs	~1×10^8 (100 million)	~5×10^15 (5 quadrillion)	~6×10^17 (600 quadrillion)	~1×10^20 (100 quintillion)	~5×10^22 (50 sextillion)

Table 1: Rough breakdown of **peak FLOPs capacity** added in each year by architecture. The percentages in parentheses indicate that category's share of the year's total. (1984 values are smaller and more uncertain; by 2024, values are enormous – e.g. 5×10^22 FLOPs is 50 *sextillion* FLOPs.) These estimates are based on historical sales volumes and typical FLOPs per device in each category, corroborated by sources where available (e.g. 1984 PC vs. mainframe sales , ARM chip shipment records , and AI processor market shares). The overall growth is staggering – over 30+ years, annual shipped computing power grew by about **a factor of 10^14** (100 trillion-fold).

Projections: 2034 and 2044

Looking ahead, projecting computing capacity a decade or two into the future involves uncertainty. However, barring any catastrophic slowdown, **Moore's Law**-style improvements (though slowing in per-transistor terms) combined with increasing demand suggest continued exponential growth in total FLOPs sold. Below we outline possible scenarios for **2034** and **2044**:

- **2034 Projection:** By 2034, it is plausible that annual new hardware could provide on the order of **10^24 FLOPs** (a **yottaFLOP** scale) of capacity. This assumes roughly a 10× to 20× increase from 2024 (equivalent to ~12–26% compound annual growth in FLOPs, slower than historical 30–60% but accounting for maturation). In 2034, **GPUs/AI accelerators** will likely account for an even larger share of FLOPs – potentially well over half. **NVIDIA** is poised to remain a key player if it continues technological leadership, but by 2034 we may see more competition from specialized AI chips (e.g. from cloud providers or startups) and possibly from **AI-optimized ARM/RISC-V designs**. We also expect **ARM/RISC-V** to solidify their dominance in general-purpose computing. By 2034, *virtually all* consumer devices and a large fraction of enterprise systems could be ARM/RISC-V based. Even the PC/server world might shift: for instance, ARM-based servers (or RISC-V servers) could seriously compete with x86 for cloud workloads. **x86** might shrink to a legacy role or a smaller high-end niche (perhaps <10% of FLOPs by then). Also by 2034, the billions of tiny microcontrollers shipping annually (often RISC-V by that time, given trends) will collectively contribute substantial computing (though much of it idle or low-utilization). **Quantum computing** may still not contribute to FLOPs (it solves certain problems differently, and any quantum "advantage" by 2034 would likely be in non-FLOP metrics). So for 2034, one can envision: **GPUs/Accelerators ~60%**, **ARM/RISC-V CPUs ~30%**, **x86 CPUs ~5%**, **Other ~5%** of the FLOPs. The total could be around **1×10^24 FLOPs** (with uncertainty, perhaps between 5×10^23 and 2×10^24).
- **2044 Projection:** Twenty years out, forecasting is even more speculative. If trends hold, 2044's new hardware might deliver **10^25–10^26 FLOPs** (ten to one hundred yottaFLOPs). By 2044, we may approach physical and economic limits of conventional silicon scaling, but by then new technologies (like **3D stacking, optical/photonic computing, or even neuromorphic and quantum hybrids**) could augment continued growth. The **architecture mix** in 2044 will likely be very different from 2024: it's conceivable that *GPUs as we know them* converge with other processors – for example, general CPUs might all be highly parallel "XPU" designs that include CPU+GPU+AI cores on one package . Still, for consistency, if we categorize by lineage: **NVIDIA (or its successors)** might still lead in high-performance parallel processing, but we could also see strong **AI accelerators from many vendors** (possibly making the "non-NVIDIA GPU" share larger than NVIDIA's). **ARM/RISC-V** (or similar open architectures) will likely comprise almost all general-purpose processors; by 2044, x86 could be largely phased out (except perhaps in legacy support or retro niche). We also anticipate **ubiquitous computing** – by 2044, tens of billions of devices from household appliances to infrastructure will have significant on-board AI compute. The FLOPs from these embedded AI chips (many using RISC-V or ARM cores with neural accelerators) will form a large portion of the total. Another factor by 2044: **cloud computing** and **edge computing** might involve a vast distributed network of accelerators – effectively, computing power will be everywhere and often specialized. If one assigns rough percentages: maybe **70% accelerators (GPUs/AI chips)**, **29% other CPUs (ARM/RISC-V)**, and **~1% legacy (if any)**. Quantum computers, if operational at scale by then, won't be measured in FLOPs (and are unlikely to replace classical computing broadly).

To summarize the projection, Table 2 provides a possible breakdown for 2034 and 2044:

Table 2. Projected FLOPs Sold in 2034 and 2044 (by architecture)

Category	2034 (proj.)	2044 (proj.)
x86 CPUs	~5×10^23 FLOPs (~5%)	~1×10^25 FLOPs (~1%)
ARM/RISC-V CPUs	~3×10^24 FLOPs (~30%)	~3×10^25 FLOPs (~29%)
GPUs/AI – NVIDIA	~6×10^24 FLOPs (~60%)	~7×10^25 FLOPs (~70%)
GPUs/AI – Others	~1×10^24 FLOPs (~10%)	~2×10^25 FLOPs (~20%)
Other (legacy/analog)	~few×10^22 FLOPs (<1%)	~few×10^23 FLOPs (<1%)
Total	**~1×10^25 FLOPs** (10^25)	**~1×10^26 FLOPs** (10^26)

Table 2: **Projected** annual compute capacity shipments for 2034 and 2044. Values are highly approximate. We assume continued exponential growth but somewhat moderated from historical 50%+ per year to perhaps 20–30% per year. By 2034, total new FLOPs per year may reach ~10^25, and by 2044 on the order of 10^26. **Architectural shares** are expected to shift further toward accelerators. We anticipate **x86** to dwindle to only a few percent or less, **ARM/RISC-V** (and other open architectures) to underpin essentially all general-purpose computing (around one-third of FLOPs), and **GPUs/AI accelerators** to contribute the majority (two-thirds or more). NVIDIA is used here as a placeholder for the leading GPU vendor – it could well still be NVIDIA, given its strong ecosystem, but it could also represent any dominant player in 2044's accelerator market.

Conclusion

Between 1984 and 2024, the world went from selling millions of FLOPs to sextillions of FLOPs each year – an astronomical increase of 14 orders of magnitude. The architectural landscape transformed from one ruled by large centralized mainframes and minicomputers to one that is richly heterogeneous. In 1984, **mainframes and minis** still delivered a large fraction of computing power, but by the 1990s the **x86 PC** had taken over. The 2000s and 2010s saw the ascent of **ARM**, riding the wave of mobile and embedded computing, and the resurgence of parallel computing through **GPUs**. As of 2024, most FLOPs are produced by specialized accelerators and billions of lightweight cores, rather than traditional CPUs alone.

Looking ahead, this trend will likely continue: **computing everywhere** means that the notion of FLOPs will permeate all devices, and **AI accelerators** will push the frontier of performance. By 2034 and 2044, we expect the total FLOPs sold annually to reach zetta- and yotta-scale, respectively, with the vast majority coming from non-x86 architectures – primarily massively parallel processors and ubiquitous embedded cores. These projections assume continued demand for computation (driven by AI, data analysis, automation, etc.) and the industry's ability to innovate beyond the limits of current silicon technology.

In summary, the evolution of FLOPs sold by architecture can be seen as a series of **paradigm shifts**: from mainframes → to PCs (x86) → to mobile (ARM) → to accelerators (GPUs/AI). Each shift added orders of magnitude more computing capacity and democratized access to compute in new ways. The numbers are staggering, but they highlight how far technology has progressed – and give a glimpse of a future where computing power is even more abundant and deeply integrated into everything around us.

Sources:

- Hilbert, M. & López, P. (2012). *"The World's Technological Capacity to Compute Information."* Science/Intl. Journal of Communication – provided global compute estimates in MIPS (e.g. ~2×10^20 IPS in 2007) .
- Stanley, J. (2018). *"What would the total computing power of the entire world in 1970 cost today?"* – blog summarizing Hilbert's data (shows growth from 10^8 MIPS in early 1980s to 10^12+ by 2000s) .
- **1984 PC vs Mainframe:** In 1984, PC sales ($11.6B) surpassed mainframe sales ($11.4B) , indicating the shifting balance of computing.
- **ARM Shipments:** ARM Holdings reports show >6.7 billion ARM-based chips shipped in one quarter of 2020 (Q4), outselling x86 and others combined . This underscores ARM's volume dominance by the 2010s.
- **GPU/AI Share:** By 2020, NVIDIA held ~80% of the cloud/data center AI processor market , reflecting the dominance of NVIDIA GPUs in high-performance computing. Also, most new Top500 supercomputer performance in late 2010s came from GPUs .
- IDC and industry reports on PC shipments, server revenues, etc., for context on market sizes (e.g. Unix server revenue ~$25B in 1999 vs. shrinking by 2012) .
- Manufacturer specifications for representative chips (Intel/AMD CPUs, NVIDIA/AMD GPUs, etc.) were used to estimate per-unit FLOPs for each era.

Wikipedia (Mainframe computer) – In 1984 estimated sales of desktop computers ($11.6 billion) exceeded mainframe computers ($11.4 billion) for the first time… (Signifying the crossover point of PC vs mainframe dominance in the market.)

Tom's Hardware (2021) – "…in the fourth quarter of 2020 alone, the Arm ecosystem shipped a record 6.7 billion Arm-based chips… This means that Arm outsells all other CPU architectures – x86, ARC, Power, and MIPS – combined." (Illustrates the massive volume of ARM chips, showing how by the 2020s ARM eclipsed x86 in unit count.)

Omdia (2021) – "Leading GPU supplier NVIDIA… [had] an 80.6% share of global revenue [for] cloud and data center AI processors in 2020." (Demonstrates NVIDIA's dominance in the AI accelerator market, a proxy for its share of shipped FLOPs in that segment.)

AI Impacts (2015) – "Hilbert & Lopez (2012) estimated that in 2007 the world's computing capacity was around 2×10^{20} IPS…Growth is unlikely to have been negative since 2007…we revise our estimate to $2\times10^{20} – 1.5\times10^{21}$ FLOPS for the end of 2015." (Gives an order-of-magnitude sense of total global computing, aligning with our 2014 estimate of ~10^{20} FLOPs added per year.)