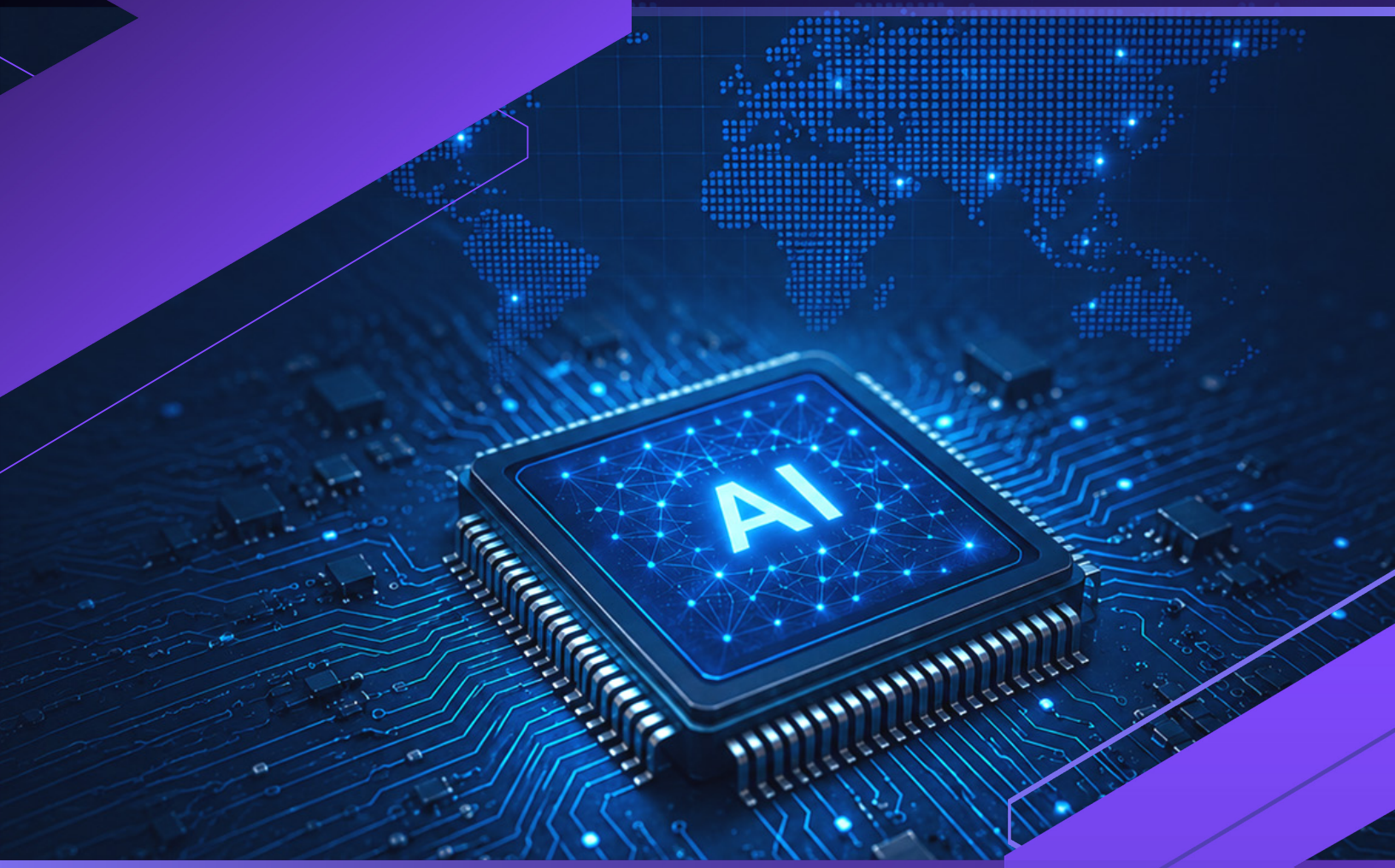
ANAQUA®




# Semiconductor Industry Patent Report

## AI & Innovation Trends

September 2026

anaqua.com

CONTENTS



Integrated Circuits (IC) are the engines of the modern technology industry, providing the computational foundation on which every digital innovation is built. Today, artificial intelligence is driving unprecedented leaps in both technology and business, transforming industries and redefining what is possible at a remarkable pace. That transformation runs on silicon.

The AI revolution was sparked by software innovations – the transformer architecture and the scaling laws that proved bigger models trained on more data yield better results. This breakthrough created an almost limitless demand for computing power, and hardware makers are moving to meet it. The result is an arms race in physical infrastructure, with semiconductors being chief among them. Software set the destination and hardware is building the road to get there.

The convergence of semiconductor and AI development is clearly visible when examined with an intellectual property lens, where it is generating a wave of patent activity that reflects the race to own the foundational technologies of tomorrow. This report examines what is happening at that intersection, exploring how AI is shaping semiconductor innovation and how that innovation is being protected and commercialized through patenting.

As a global leader in intellectual property management systems and services, Anaqua is uniquely positioned to uncover insights of this rapidly evolving landscape. Leveraging its advanced patent information search capabilities, Anaqua undertook a deep dive into the state of semiconductor and AI patenting, drawing on a rich body of patent data to identify trends, surface key players and map the contours of innovation at this critical intersection. The findings presented in this report reflect that research, offering IP professionals, business leaders and technology strategists a clearer picture of where the industry has been and where it is heading.

# Highlights (Based on a five-year analysis from 2021-2026)

**AI is a disproportionate, structural driver of semiconductor innovation – it's not a cyclical bump.** AI x Semiconductor patents grew 114% versus 78% for the broader combined universe. A closer examination of key semiconductor subcategories across the AI landscape makes it even more evident that artificial intelligence is driving an outsized impact on the chip market. This isn't a niche trend: AI has become the primary organizing force behind where semiconductor R&D dollars and patent filings are being directed.

**Samsung is leveraging its memory position into broad AI-hardware leadership.** Samsung leads AI architecture (1,194 filings), leads HBM (107), and ranks second in inference and analog AI – spanning neural processing units (NPU), processing-in-memory, and Exynos system-on-a-chip (SoC). It's the clearest case of a memory manufacturer converting that base into end-to-end AI-chip strength.

**NVIDIA's patent footprint may appear smaller relative to its market dominance – because its moat isn't in hardware.** NVIDIA holds just 49 GPU-titled semiconductor patents (#7) and ranks #15 in inference chips. Its durable advantage lies in the CUDA software ecosystem (introduced in 2006).

**China dominates AI inference patenting by a significant margin and is rising fast in GPUs.** Chinese government-affiliated entities lead inference/accelerator filings with 5,387 over five years – far ahead of Samsung's 1,897 – spread across universities, national labs and state research institutes. The same entities rank third in graphic processing unit (GPU) filings, with startups like Moore Threads and Muxi emerging. This can be read as the result of a chip-sovereignty response to U.S. export restrictions.

**Intel is strategically positioned for an AI future.** Intel leads GPU semiconductor filings with 188. It is second in patents, with 967, in the AI in the IC architectural space. Intel ranks third with 23 patents that address HBM. It is fifth in the number of patents granted (225) at the convergence of AI and analog ICs. Additionally, it ranks 8th in inference semiconductor patents with 792.

**IBM leads where the science is hardest.** It tops AI-semiconductor crossover filings (794) and analog AI (389), with a footprint concentrated in frontier technologies. These include quantum, neuromorphic chips and phase-change memory that encodes neural-network weights directly onto the chip (claiming ~14x better energy efficiency, now reaching commercial deployment via its Spyre accelerator). Its declining inference filings (-31%) suggest a deliberate focus on differentiated, long-horizon bets over conventional accelerator volume.

**Qualcomm is pivoting from mobile into the data center, fast.** Long anchored in on-device/edge AI, Qualcomm now ranks #9 in inference (655 filings, notably with none granted five years ago) and shows some of the steepest growth in the report – +186% in AI architecture (252 filings) and a rising GPU position (#4, 134 filings). Its analog work spans neuromorphic mobile circuits and its new AI200/AI250 data-center inference accelerators, which lean on power efficiency and memory capacity as the differentiators against competitors.

**Hyperscalers have become silicon designers.** Alphabet/Google's tensor processing unit (TPU) and Microsoft (Maia accelerators)-related patenting appear across nearly every AI and semiconductor patenting intersection. This reflects cloud giants designing custom chips for their own workloads rather than buying off the shelf – a vertical-integration shift emerging in terms of who controls AI hardware.

**Analog/processing in-memory (PIM) computing is an under-appreciated frontier defined by radically divergent bets.** With more than 5,700 patent filings, this subgroup features fundamentally different conceptions of computing. IBM's phase-change memory (claiming ~14x efficiency), Microsoft's optical computing, Intel's neuromorphic Loihi and Samsung's MRAM-based PIM are some key examples. Competition here runs not just between companies but between architectures and nations – looking to tackle the von Neumann data-movement bottleneck that constrains all digital AI.

# The Overall Semiconductor Market [1]

Annual publication volumes evolved across five consecutive 12-month windows. Growth was modest between the first and second year (+18%), then surged in the middle period (+29%) before tapering in the most recent windows. This pattern is consistent with the lag between R&D investment decisions – triggered by chip shortages and AI demand starting in 2021–2022 – and their appearance as published patents 18–30 months later.

**713,755**
5-YEAR TOTAL

**174,048**
LATEST 12 MONTHS

**+78%**
5-YEAR GROWTH

| PERIOD | '21–22 | '22–23 | '23–24 | '24–25 | '25–26 |
|---|---|---|---|---|---|
| Annual Filing Count | 98,039 | 114,381 | 155,835 | 171,452 | 174,048 |

The sharpest acceleration occurred in the Y-3 to Y-2 window (+29% YoY), corresponding to filing decisions made during 2021–2022. This coincides with the global chip shortages peak, where governments launched major subsidy programs (CHIPS Act, EU Chips Act, China's Big Fund III). The recent flattening (+2%) may signal the market reaching a new steady-state, or may reflect publication backlogs at patent offices that are processing record volumes.

# AI & Semiconductor Convergence [2]

To roughly measure AI's impact on semiconductor patenting, a quick search found that patents claiming both AI and machine Learning (ML), and semiconductors found 9,101 patents over a five-ear window. The year-over-year trend of patents (between 2021-May of 2026) claiming both AI/ML methods and semiconductor innovations grew +114% from 1,064 to 2,274 annually – significantly outpacing the 78% growth of the broader combined universe, confirming AI is a significant driver of semiconductor innovation.

## AI × Semiconductor Filing Trend

| PERIOD | '21–22 | '22–23 | '23–24 | '24–25 | '25–26 |
|---|---|---|---|---|---|
| AI × Semi Annual Count | 1,064 | 1,446 | 2,090 | 2,227 | 2,274 |

The overall stability of the most recent rankings suggests these AI-semiconductor investment patterns are structural, not cyclical. The intersection is now a permanent and growing feature of the semiconductor patent landscape.

A closer analysis of key semiconductor subcategories across the AI landscape makes it even more evident that artificial intelligence is driving an outsized impact on the chip market.

**9,101**
AI × SEMI (5-YEAR)

**+114%**
AI × SEMI GROWTH

[1]A methodological note: The universe reported in Section 1 combines all search pools used in this report: the base semiconductor scope (CPCF: H01L/H10), GPU patents (Title (TTL) TTL: GPU or graphics processing unit, filtered to G06F/G06T/H01L/H10), inference/accelerator patents (TTL: accelerator, inference, tensor, neural network, deep learning, intersected with CPCI:G06N), AI architecture patents (CPCI:G06N intersected with G06F15/G06F7/G06F9), and analog AI patents (CPCI:H03F/H03M/H03K intersected with G06N). All results are deduplicated at the application family level using AFAM_REP:T. Because these search pools use different CPC classifications – many AI and GPU patents are classified under G06F/G06N rather than H01L/H10 – the combined universe is larger than the base semiconductor scope alone.
The Cooperative Patent Classification (CPC) underwent a major reclassification that moved most semiconductor device patents from H01L into the new H10 class. Any analysis relying solely on H01L will capture only ~3,000 of the actual ~612,000 universe—a 99.5% undercount. This is essential context for interpreting any semiconductor patent data.

[2]Methodological Note: The search intersected the CPC class G06N (AI and machine learning) with semiconductor classes (H01L/H10) to find patents claiming both AI methods and semiconductor innovations.

# GPU Patent Landscape

Graphic processing units (GPU) are versatile, high-throughput parallel processors. They were originally designed for, and are still used for, graphics processing. However, GPUs have become the enabling technology for AI. GPUs are flexible, parallel processors that have become the industry standard for both training AI models and running inference.

In GPU-titled semiconductor patents, Intel leads all filers with 188 filings over the five-year period, reflecting its aggressive push into discrete GPUs with the Arc product line and data center GPU efforts. AMD ranks second (178 filings), and combined with ATI (42 filings), its effective total is larger than Intel's. Qualcomm is fourth with 134 filings over the last five years, with an initial emphasis on graphics and on-device computing, but with a more recent emphasis on data center computing.

**3,431**

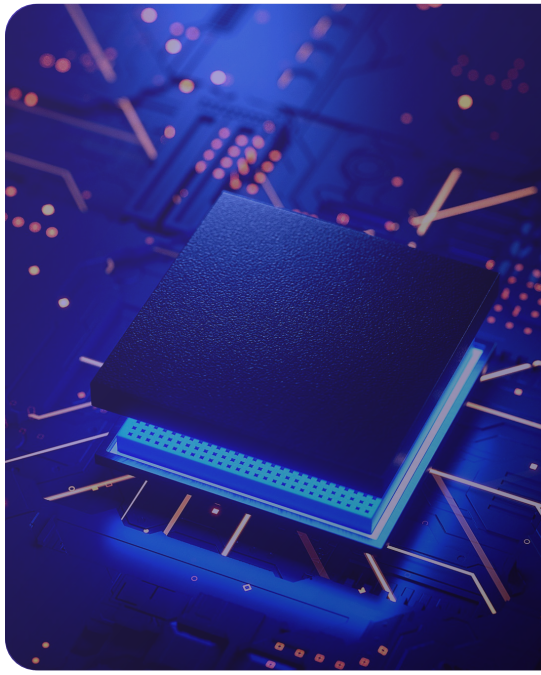
GPU TOTAL (5-YEAR)

**+108%**

GPU GROWTH

| COMPANY | 5-YEAR TOTAL | LAST 12MO | NOTES |
|---|---|---|---|
| **Intel** | 188 | 66 | Arc GPU line + data center GPU push |
| **AMD** | 178 | 39 | Combined with ATI/AMD legacy (~195 total) |
| **China Gov't Entities** | 155 | 39 | Domestic GPU alternative program |
| **Qualcomm** | 134 | 23 | Mobile / Datacenter GPU |
| **Imagination Technologies** | 107 | 29 | UK GPU IP specialist |
| **Sony** | 72 | — | Gaming + image processing |
| **NVIDIA** | 49 | 18 | GPU semiconductor-classified only |
| **Moore Threads (China)** | 23 | 14 | Chinese GPU startup (est. 2020) |

Notably, Chinese entities are increasingly active in GPU patenting: government-backed entities rank third in recent filings, with 155 over the last five years – with Inspur, Moore Threads and Muxi Integrated Circuit also filing. This likely represents China's push to develop domestic GPU alternatives in response to U.S. export restrictions on advanced AI chips. Moore Threads, a Chinese GPU startup founded in 2020, has risen to seventh in the most recent 12-month GPU rankings with 14 filings.



NVIDIA's relatively smaller position with just 49 GPU-titled semiconductor patents over five years – only 18 in the most recent 12-month window – might come as a surprise given their absolute dominance in the GPU industry. NVIDIA's relatively modest count of GPU-titled patents belies the true depth of its competitive moat. The company's most durable advantage lies not in GPU patents alone. Its Compute Unified Device Architecture (CUDA), the proprietary parallel computing platform and programming model it introduced in 2006, transformed NVIDIA's GPUs from graphics accelerators into general-purpose computing engines, enabling developers to harness GPU parallelism for scientific simulation, deep learning and data analytics.

A methodological note: GPU patents were identified by searching for patents mentioning 'GPU' or 'graphics processing unit' in the title, filtered to relevant CPC classes (G06F computing, G06T image processing, H01L/H10 semiconductors). This captures patents focused on GPU hardware and GPU-specific computing architectures.

# Inference / Accelerator Semiconductors

The AI industry has been shifting focus from training LLMs to operating them at scale. This computational phase is known as inference. While GPUs do perform inference calculations, this versatility comes at the cost of high power consumption and heat. In contrast, an inference chip is a specialized "AI accelerator" built strictly for inference processing. By stripping away non-essential hardware used for gaming or 3D rendering, the inference chips focused entirely on the mathematical operations that power LLM neural networks, delivering similar performance to a GPU while using significantly less energy.

This section of the report is primarily concerned with specialized AI-inference patents for chips. Additionally, beyond standalone integrated circuits (IC), there are companies that offer AI-related inference IP cores that reside in processors, application specific integrated circuits (ASIC), systems-on-a-chip (SoC) and in some cases within networking/ communications modules related to data movement efficiency.

China government entities lead by a wide margin at 5,387 five-year filings. This reflects the breadth of state-directed AI hardware investment across universities, national labs, and government-affiliated research institutes. Zhejiang Univ (1,247), State Grid of China (802), Chongqing University (600) and Southeast University (506) all rank highly on their own among large inference AI semiconductor patent holders. Additionally, Huawei (973) ranks fifth.

| COMPANY | 5-YEAR TOTAL | LATEST 12MO | GROWTH |
|---|---|---|---|
| **China gov't entities** | 5,387 | 1,691 | +26% |
| **Samsung Group** | 1,897 | 701 | +52% |
| **Zhejiang Univ** | 1,247 | 362 | -12% |
| **Alphabet / Google** | 1,087 | 330 | +16% |
| **Huawei** | 973 | 289 | +5% |
| **Intel** | 825 | 291 | +28% |
| **State Grid China** | 802 | 254 | +28% |
| **IBM** | 792 | 195 | -31% |
| **Qualcomm** | 655 | 234 | — |
| **Korea gov't entities** | 606 | 200 | +14% |
| **Chongqing Univ** | 600 | 212 | +33% |
| **Microsoft** | 554 | 148 | -17% |
| **Bosch** | 541 | 164 | +1% |
| **Southeast Univ** | 506 | 158 | -2% |
| **NVIDIA** | 501 | 189 | +41% |

Samsung ranks second with 1,897 filings over five years. Samsung is more focused on NPU-related technologies. While these have been more focused on edge computing, their speed and capabilities have made them a fast growing segment of data center computing technology.

Ranking fourth is Alphabet/Google (1,087) and at 12th is Microsoft (554). The hyperscalers are making significant investments in inference chip innovation to meet their needs.

Intel (825) ranks sixth, demonstrating a strong investment in inference computing. IBM (792) ranks eighth which shows that they have made significant investments in AI inference processing.

Qualcomm (655) ranks ninth which is interesting if you consider that most of its patent grants in inference semiconductors have taken place over the last few years, with none granted as of five years ago.

Bosch (541) ranks 13th and is more focused on patenting inference-related technologies for the edge, specifically within sensors to process data where it is collected. Rounding out the Top-15 holder of AI inference semiconductor patents in NVIDIA (501). It provides data processing units (DPU) and its NVIDIA Deep Learning Accelerator (NVDLA).

A methodological note: To capture this data, the analysis intersected "accelerator" with (AI/ML), TTL:(inference OR tensor OR "neural network" OR "deep learning" OR accelerator OR NPU OR "forward pass" OR "model execution" OR "tensor streaming" OR "deterministic streaming") AND CPCI:G06N AND AFAM_REP:T

# AI at the IC Architecture Level

A critical layer – where many semiconductor companies appreciably innovate – is the IC architecture level. The architectural design phase is the foundational planning stage of integrated circuit development, where engineers define what a chip must do before determining how it will be built.

At this level, designers map out functional block diagrams, specify input/output pins and establish intellectual property (IP) blocks. They also devise strategies for clocking and reset, all while keeping performance and power management targets front of mind.

**15,119**
AI ARCH (5-YEAR)

**3,465**
AI ARCH (12 MO)

**+54%**
5-YEAR GROWTH

By resolving these high-level decisions early, the architectural phase ensures the design is oriented toward its core goals, whether that's energy efficiency, processing speed, or both, before the far more labor-intensive work of logic development begins: processor design, parallel computing structures, arithmetic unit organization, instruction pipeline architecture and hardware power management.

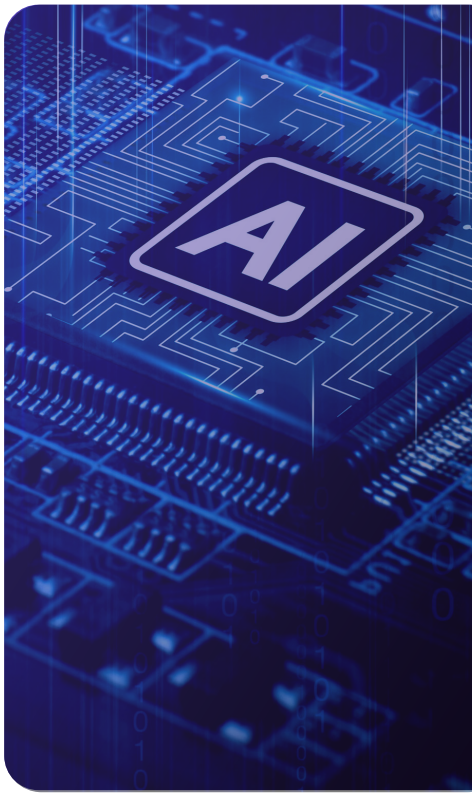
Analysis of patents combining AI/ML methods with architecture innovations that include processor design, parallel computing structures, arithmetic unit organization, instruction pipeline architecture and hardware power management, reveals a landscape that looks different from either the semiconductor-device or GPU-titled rankings.

## Five-Year Rankings

| COMPANY | 5-YEAR TOTAL | LAST 12MO | CHANGE | NOTES |
|---|---|---|---|---|
| **Samsung** | 1,194 | 356 | **+176%** | NPU + PIM investment paying off |
| **Intel** | 967 | 190 | **-2%** | Lost architecture leadership position |
| **Alphabet/Google** | 736 | 141 | **+36%** | Steady TPU evolution |
| **IBM** | 382 | 65 | — | Quantum, neuromorphic |
| **Microsoft** | 328 | 52 | **-35%** | Pivot from build to partner? |
| **NVIDIA** | 252 | 57 | **+82%** | Growing in line with market |
| **Huawei** | 253 | 58 | **+152%** | Accelerating despite sanctions |
| **Qualcomm** | 252 | 48 | **+186%** | Aggressive mobile AI push |
| **Cambricon (China)** | 200 | — | declined | China's leading AI chip co. |
| **Fujitsu** | 202 | — | — | Japanese AI processor research |
| **Micron** | 195 | 24 | — | Memory-compute convergence |
| **TSMC** | 148 | 62 | new | Foundry co-design trend |
| **Rebellions (Korea)** | 126 | 52 | new | Korean AI chip startup |
| **DeepX (Korea)** | 105 | 39 | new | Korean AI chip startup |
| **Graphcore (UK)** | 95 | — | — | UK AI processor specialist |

A methodological note: To capture this layer, the analysis intersected G06N (AI/ML) with architecture-level CPC codes: G06F15 (processor architecture, array processors), G06F7 (arithmetic/logic units), G06F9/30 (instruction processing, pipelining), and G06F1 (hardware details including power management). This identifies patents claiming both AI methods and processor architecture innovations—the design of AI-optimized chips at the circuit and microarchitecture level.

## AI at the IC Architecture Level (cont.)




Samsung leads AI architecture patenting with 1,194 filings over five years, reflecting NPU development for Exynos mobile system-on-chips (SoC), its Mach-1 AI accelerator and processing-in-memory research. Intel ranks second with 967 filings, redirecting decades of processor architecture leadership toward AI acceleration. Alphabet/Google ranks third at 736 filings, with architecture patents spanning TPU systolic array designs, custom interconnects and AI-specific memory hierarchies.

NVIDIA grew 82% at the architecture level, from 28 to 57 filings in the last 12 months, outpacing the 54% growth of the AI architecture market overall. It ranks sixth at the architecture level with 252 filings over the last five years.

In the most recent 12-month window, Samsung extended its architecture led to 356 filings, up 176% from five years ago. Intel held essentially flat at 190, a -11% change in a market that grew 54%.

Huawei ranks 7th with 253 filings over the last five years and Cambricon, China's leading AI chip architecture company, ranks ninth with 200 filings.

Korean startups Rebellions (126 filings) and DeepX (105 filings) continue to file at rates approaching AMD (144 filings) and Apple (89 filings).

Graphcore (95 filings) from the UK adds rare European representation.

## Analog × AI: An Emerging Frontier.

The analog and analog-adjacent AI chip sector represents one of the most important – and underappreciated – frontiers in semiconductor innovation. Unlike conventional digital chips, which shuttle data back and forth between separate memory and processing units at enormous energy cost, analog AI chips perform computation directly where data is captured and stored. This "in-memory computing" approach mimics the efficiency of biological neural networks, and the potential gains are substantial.

**5,732**

ANALOG × AI (5-YEAR)

**~4%**

SHARE OF ANALOG

| COMPANY | 5-YEAR TOTAL | FOCUS |
|---|---|---|
| **IBM** | 389 | Neuromorphic computing, analog AI accelerators |
| **Samsung** | 285 | AI-optimized signal processing |
| **Microsoft** | 277 | Quantum computing (analog-adjacent) |
| **Alphabet/Google** | 226 | AI-driven signal processing |
| **Intel** | 225 | AI hardware design |
| **Huawei** | 176 | RF and signal AI optimization |
| **Atombeam Tech** | 139 | Data compression / transmission AI |
| **Qualcomm** | 124 | Mobile AI signal chain |
| **Northrop Grumman** | 101 | Defense/aerospace edge AI inference |

# Analog × AI: An Emerging Frontier (cont.)

The top patent holders reflect three distinct technical approaches. IBM, the leader, uses phase-change memory (PCM) to encode neural network weights directly onto analog chips. It has demonstrated up to 14 times better energy efficiency than digital counterparts, with its Spyre accelerator now moving toward commercial deployment in enterprise mainframes.

Samsung, second in filings, leverages its position as the world's largest memory manufacturer to pioneer MRAM-based in-memory computing and Processing-in-Memory (PIM) in High Bandwidth Memory stacks.

Microsoft is pursuing the most radical approach. Its Analog Optical Computer (AOC) uses light rather than electricity to perform computation, combining microLED arrays and spatial light modulators to achieve a projected 500 tera-operations per second per watt. This is potentially 100 times more efficient than leading GPUs.

Intel's Loihi neuromorphic processor mimics the actual firing behavior of biological neurons using spiking neural networks that consume just 100 milliwatts – targeting edge AI in healthcare, robotics and autonomous vehicles with commercial rollout planned for Q3 2026.

Google's TPU program, while not purely analog, attacks the same von Neumann bottleneck through systolic array architecture; its eighth-generation TPU 8t and TPU 8i chips deliver up to 2x better performance-per-watt than their predecessors and scale to 121 ExaFlops per superpod.

Huawei's 176 filings reflect a parallel Chinese development path, centered on neuromorphic algorithms at its Zurich Research Center and event-driven edge processing for IoT and 5G.

Qualcomm's 124 patents span neuromorphic mobile circuits and its new AI200/AI250 data center inference accelerators, which emphasize power efficiency and high memory capacity as key differentiators against NVIDIA.

The final two entrants reveal how far beyond traditional chipmakers this technology has spread. Atombeam Technologies, with a surprising 139 patents for a startup, takes a software-layer approach – its Data-as-Codewords technology uses machine learning to compress IoT and machine-generated data.

This is a sector whose competitive lines run not just between companies but between nations, architectures and entirely different conceptions of what a computer can do.

Methodological Note: Intersecting analog circuit CPC codes (H03F/H03M/H03K) with AI/ML codes (G06N) identified 5,732 patents.

## About the Report

Anaqua's Semiconductor Industry Patent Report: AI & Innovation Trends is based on an analysis of total semiconductor-related filings over the latest five-years period. This report was generated using data from Anaqua's AcclaimIP patent analytics platform, accessed in between April and May of 2026. All queries were executed via the AcclaimIP MCP (Model Context Protocol) connector, enabling direct programmatic access to the full patent database.

## About Anaqua, Inc.

Anaqua is a premier provider of integrated intellectual property (IP) management technology solutions and services for corporations and law firms. Its IP management software platforms, AQX®, PATTSY WAVE®, and RightHub® offer best practice workflows with big data analytics and tech-enabled services to create an intelligent environment designed to inform IP strategy, enable IP decision-making and streamline IP operations – tailored to each segment's needs. Today, nearly half of the top 100 U.S. patent filers and global brands, as well as a growing number of law firms worldwide use Anaqua's solutions. Over two million IP executives, attorneys, paralegals, administrators and innovators use Anaqua's platforms for their IP management needs. The company's global operations are headquartered in Boston, with offices across the U.S., Europe, Asia and Australia.



ANAQUA®

U.S HEADQUARTERS
31 St. James Ave Suite 1100
Boston, MA 02116, USA
+1 617 375 5808

OTHER OFFICES
Australia | France | Germany | India | Italy | Japan | United Kingdom