A Brief History of VSORA

VSORA traces its origins to DiBcom, a company founded in 2000 by Khaled Maalej and other cofounders, now VSORA's CEO. DiBcom specialized in digital mobile TV receivers and rapidly became the global leader in its niche—at one point holding nearly 100% market share. After its acquisition in 2012, the leadership team dispersed to pursue new ventures.

The seeds of VSORA were planted in 2015. During their time at DiBcom, the team had worked extensively with major European automakers to integrate mobile TV technology into vehicles. While their designs were tailored to European broadcast standards, two major German automotive manufacturers eventually approached them with a new challenge: to develop a single, reconfigurable chip capable of operating across all global digital TV standards—covering Europe, the U.S., China, and Japan.

Encouraged by this opportunity and backed by strong customer commitment, the team set out to license a suitable DSP (Digital Signal Processor). However, no existing DSP could meet the extreme computational demands of the proposed universal TV receiver. As a result, they were forced to design their own DSP—along with a full compiler toolchain—an effort they later described as a "nightmare." Although the chip ultimately reached production and remained in use for years, the experience reinforced a critical lesson: future architectures had to be built on widely adopted industry standards.

By 2013, LLVM had emerged as the open-source compiler with the highest potential. The team resolved that any new DSP architecture would leverage such standardized technologies. Equally important, they decided not to re-enter the semiconductor manufacturing business itself, opting instead for an IP licensing model to avoid the heavy cost and complexity of running a fab operation.

Their first new project was a 5G software-defined radio (SDR) receiver. During customer outreach, they reconnected with long-standing automotive partners, several of whom suggested exploring artificial intelligence. These partners observed that the architecture underlying the team's DSP closely mirrored the computation model used in AI accelerators. A deeper analysis confirmed this: more than 90% of the architectural components overlapped.

Within months, VSORA launched a new AI-focused product line. A major German automaker soon proposed combining DSP and AI capabilities into a single platform for autonomous driving. Their motivation was twofold: performance needs for autonomous systems and regulatory concerns. While AI systems can exhibit non-deterministic

behavior, DSP algorithms are mathematically verifiable, providing a defensible foundation in the event of legal scrutiny following automotive accidents.

After 18 months of collaboration, the team delivered a processor IP core in 2020 that achieved one petaflop of performance—vastly exceeding competitive offerings at the time, including NVIDIA's Orin platform, which delivered 128 teraflops. The Linley Group recognized this achievement by naming it Processor IP of the Year 2020.

As discussions continued, the team acknowledged the significant market opportunity and chose to re-enter semiconductor design—this time as a fabless company. They initiated development of a new chip, the AD1028, targeting TSMC's 7nm process.

By late 2022, the rapid rise of generative AI—accelerated by OpenAI's breakthroughs—reshaped industry priorities. Customers began asking whether VSORA's architecture could support Generative AI (GenAI) workloads. A 2022 McKinsey analysis estimating Google's cost per generative query at scale ($0.002 at 100,000 QPS) became a new benchmark for cost-effective inference.

Exploring this new direction, the team concluded that the dominant cost driver for GenAI was not only power consumption but mainly the capital expenditure required for large-scale compute infrastructure. VSORA's architecture demonstrated an important advantage: it scaled efficiently with model size and maintained near-theoretical performance across deployment environments, making it particularly well-suited for large language model (LLM) inference.

Building on the AD1028 foundation, VSORA began transitioning toward a chiplet-based design aimed at two markets simultaneously: autonomous driving and data center inference. They intentionally avoided the AI training market, recognizing NVIDIA's overwhelming dominance in that segment.

Work continued on the 7nm version until mid-2023, when a visit from TSMC prompted a strategic shift to the 5nm node and an increase in compute performance to 3.2 petaflops.

The chip is now on the manufacturing schedule, with first silicon expected in Q1 2026. Supporting boards and rack-scale platforms are also underway. By the close of Q1 2026, full benchmarking will confirm—on hardware—the performance projected through two years of rigorous simulation, emulation, and FPGA prototyping.