FROST & SULLIVAN

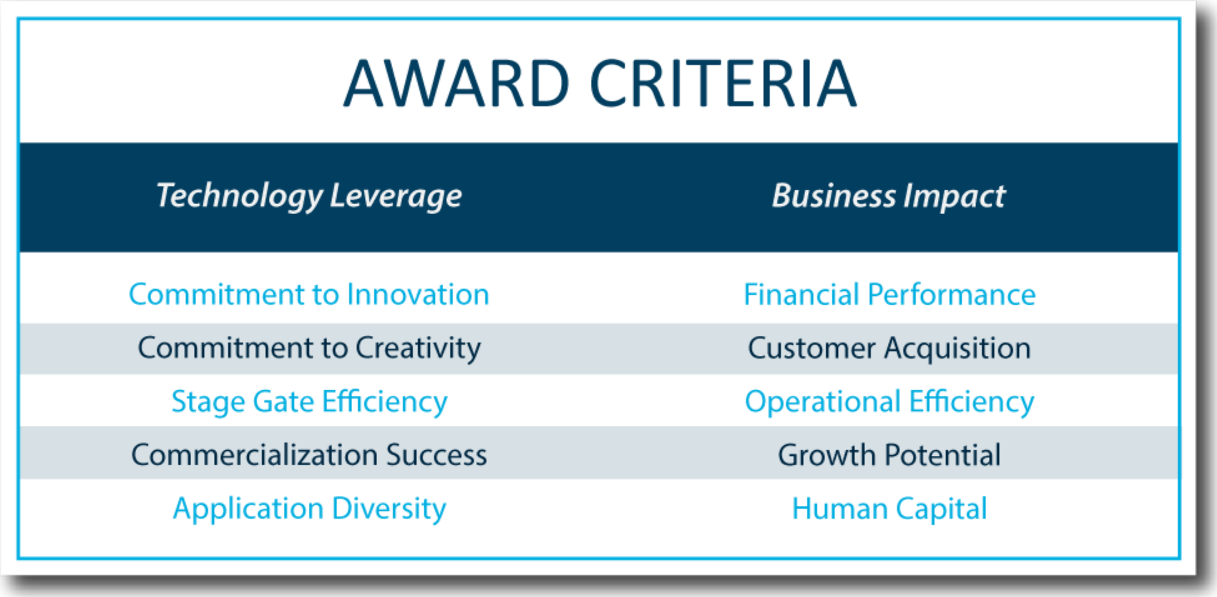
groq™

2022 TECHNOLOGY INNOVATION LEADER

North American AI Processors for Data Centers Industry

Best Practices Criteria for World-Class Performance

Frost & Sullivan applies a rigorous analytical process to evaluate multiple nominees for each Award category before determining the final Award recipient. The process involves a detailed evaluation of best practices criteria across two dimensions for each nominated company. Groq excels in many of the criteria in the AI processors for data centers space.



AWARD CRITERIA

Technology Leverage	Business Impact
Commitment to Innovation	Financial Performance
Commitment to Creativity	Customer Acquisition
Stage Gate Efficiency	Operational Efficiency
Commercialization Success	Growth Potential
Application Diversity	Human Capital

Next Generation Innovation

Founded in 2016 and headquartered in Mountain View, California, Groq is a fabless semiconductor company that develops and distributes accelerators based on the Tensor Steaming Processor (TSP) architecture for artificial intelligence (AI) and high performance computing (HPC) applications. Standard hardware architectures of the central processing unit and graphical processing unit (GPU) struggle to handle the increasing complexity of neural network algorithms. Frost & Sullivan notes that the accelerating change in AI and machine learning (ML) models and the legacy need to custom code each new model iteration presents a significant challenge, as a custom coding-based industry struggles to keep pace with the latest technology. Moreover, rising complexity stemming from hardware scaling with multiple processor cores, on-chip, modules, and threads present significant hurdles. Frost & Sullivan finds the critical step for vendors is to simplify development, which is where Groq finds itself quite well-positioned to address this complexity.

> *"Groq based its GroqChip™ processor on a deterministic instruction set architecture with a single large core, translating to functional latency. In terms of computer processing, determinism allows software to track tensors and understand a program's correctness. As a result, Groq software controls the total order of the TSP, as hardware is incapable of reordering events and their completion time."*
>
> ***- Samantha Fisher,***
> ***Best Practices Research Analyst***

The flagship Groq TSP architecture and compiler accelerates AI workloads based on complex neural network algorithms in the cloud. The TSP GroqChip™ performance is 750 tera operations per second (INT8, @ 900Mhz). For ResNet-50 version two, the chip achieves 21,700 inferences per second with ~0.5 msec latency, making GroqChip one of the fastest neural network processors currently available. The company's single-core processor possesses multiple functional units capable of handling 4,000,000 integer multiple accumulate that operates per cycle with a 230-megabyte static random-access memory (SRAM). The TSP architecture is a material departure from traditional solutions. With nearly seven times the compute density per transistor, Frost & Sullivan recognizes how Groq's approach leverages a unique combination of an easy-to-program system that enables developer velocity and offers low latency and massive throughput in a single compute core that scales near-linearly across multiple chips.

Simplifying Compute

GroqChip is based on a deterministic instruction set architecture with a single large core, translating to functional latency. In terms of computer processing, determinism allows software to track tensors and understand a program's correctness. As a result, Groq software controls the total order of GroqChip, as hardware is incapable of reordering events and their completion time. Frost & Sullivan notes that the end outcome includes reduced total cost of ownership for customers with diverse service level agreements (SLAs) and scalability for large training and inference systems. To achieve this scalability, the company leverages Reversed Scaling Unit Economics, which Groq defines as "low latency combined with high-speed switch integration and near-linear scaling." Frost & Sullivan's research finds the Groq TSP architecture unique, and its deterministic nature well-positions it as a critical vendor in applications requiring real-time latencies.

> *"Groq's TSP architecture offers optimal responsiveness along with maximum throughput resulting in superior performance. As a result, GroqChip executes complex algorithms in applications such as high energy physics and fintech."*
>
> ***- Sushrutha Katta Sadashiva, Industry Analyst***

Moreover, Groq aims to defy gravity, which, in this context, refers to the resistance to change, moving from the "We have always done it this way" mentality toward new and unfamiliar methodologies. To this end, the company delivers differentiated features (e.g., sub-millimeter low latency and predictable performance) and outperforms traditional tactics in critical areas, such as performance, power, and accuracy. Groq also simplifies and streamlines switching by eliminating the need to allocate large teams for each customer's integration and deployment, empowering them to deploy GroqChip through a simple-to-follow user interface. More importantly, while some other competing vendors optimize their devices for either latency (e.g., field-programmable gate array) or throughput (i.e., GPU), Frost & Sullivan analysts appreciate how Groq stands apart by delivering both. The company further compounds its value proposition with its ability to predictably meet extremely low latency SLAs while multiplying performance in real-time control systems. Groq demonstrated this in its work with Argonne National Laboratory when the goal was to maximize performance within a 1ms hard requirement needed for forecasting plasma instabilities in tokamak reactors. Groq exceeded this requirement at 193k IPS at 0.6ms, giving 622x higher throughput compared to the lowest latency GPU. Groq's advantage lies in the fact that the reactor's control system requires predictable execution within a specified time window for operational efficiency and safety. Groq's deterministic architecture provided zero variation in time to result. Additionally, with such high performance at low batch sizes, this enabled more complex algorithms within the required response time window.

Groq's TSP architecture offers optimal responsiveness along with maximum throughput resulting in superior performance. As a result, GroqChip executes complex algorithms in applications such as high energy physics and fintech. Frost & Sullivan observes how Groq's software-first approach, and its innovative technology drives its optimization, control, and planning functions, resulting in high performance per millimeter of silicon, saving room for computation.

Key Success through Novel Technology

Groq designed GroqChip to carry out intense AI workloads in the cloud, including a HPC cloud platform. The company has use cases that include autonomous vehicles (AV), natural language processing, and security. Groq's TSP architecture is ideal for systems when processing data in real-time applications. AV navigation, obstacle avoidance, and anomaly detection, all processing in real-time without sacrificing performance for accuracy when compared to legacy options. Additionally, GroqChip integrates into data centers dedicated to handling AI workloads, where it can operate independently or integrated to help accelerate parts of a system that take advantage of the TSP architecture. The United States houses many major high-capacity data centers, followed by China and Europe, thus providing a global opportunity for Groq to implement its technology and simultaneously position itself in the market. The company finds that latency SLAs attempt to solve problems by throwing more hardware at it when the core issue is slow responsiveness. By ensuring that customers receive both latency and throughput (instead of forcing a choice between the two) Frost & Sullivan applauds the way that Groq delivers key value and performance.

Determinism is a key selling point for Groq, as the methodology guarantees the customer what they get before buying, deploying, and using GroqChip processors. Moreover, as the customer's business grows, Groq scales the deployments and drives the cost toward zero relative to the competition. When combined with the savings associated with low switching costs and high performance across workloads, the return on investment becomes a critical differentiator. Since the initial GroqChip launch in October 2019, Groq continues to up the ante on its technology, closing the distance with the market's legacy incumbent. Groq also finds disruption comes through performance, ease-of-use, developer velocity, and deployment. To this end, it continues to simplify the digital transformation journey with its software-defined hardware that enables the software to perform operations planning - offering additional memory bandwidth and improved programming efficiency. With a Turing complete architecture, Groq continues to open different workloads to ML innovations, which well-positions it for the future.

Growth Fueled by Talent Density and Innovation

Groq approaches growth holistically, ensuring its future by dividing its attention on various fronts. The company's early focus was analytics, in which it easily differentiated itself (since broad compiler functionality was still unknown). As of 2021, low latency sensors and systems with massive real-time computer requirements (e.g., scaled physics experiment) and simulation at scale are the critical needs of the hour, with determinism playing a crucial role. Groq focuses on talent density, bringing the industry's best together to usher in the era of deterministic compute. The company maintains an innovative and collaborative culture to bring breakthrough ideas to reality, ultimately delivering rewarding experiences to customers. Groq has more than tripled its geo-agnostic team in the last year, despite the COVID-19 pandemic, enabling it to maintain high-quality service and grow its portfolio to match customer and market needs.

On the financial front, Groq's funding recently crossed $350 million, with various rounds with investors such as D1 Capital and Tiger Global. In addition, Groq recently acquired dataflow systems pioneer Maxeler Technologies, further advancing converged HPC and ML solutions by combining Groq's architecture and Maxeler's rich systems portfolio. Additionally, Groq is shipping its most recent technology, including GroqCard™ accelerators, GroqNode servers, and GroqWare™ suite – its software development kit solution, to its global customers. As a trending technology, the AI chip continues to garner key interest from global technologists and venture capitalists. The Groq TSP architecture-based chips exhibit high performance and low latency, with the ability to customize them for any AI workload; the company designed them for general purpose use cases that depend on the cloud for processing deep learning and neural network algorithms. The company enables various operators to integrate its hardware platform and accelerate AI workloads through emerging business models, thereby offering access to developers for experiments globally.

Conclusion

As AI and ML continue to evolve, the traditional method of custom coding each new model iteration presents a serious issue. Additionally, hardware scaling with various processor cores, on-chip modules, and threads is quite complex, presenting significant hurdles.

To address these issues, California-headquartered Groq developed its flagship Tensor Streaming Processor architecture and compiler to accelerate real-time AI and HPC workloads with high performance and an easy-to-program system that delivers low latency. Leveraging determinism and Reversed Scaling Unit Economics, Groq achieves a lower total cost of ownership and near-linear scalability. The company continues to advance its technology and engage in world-class partnerships. In addition, Groq focuses on talent density, ensuring its teams include the industry's best to innovate and solve customers' toughest compute challenges.

For its innovation prowess, best-in-class methodologies, experienced internal teams, and strong overall performance, Groq earns the 2022 Frost & Sullivan Technology Innovation Leadership Award in the AI processors for data centers industry.

What You Need to Know about the Technology Innovation Leadership Recognition

Frost & Sullivan's Technology Innovation Leadership Award recognizes the company that has introduced the best underlying technology for achieving remarkable product and customer success while driving future business value.

Best Practices Award Analysis

For the Technology Innovation Leadership Award, Frost & Sullivan analysts independently evaluated the criteria listed below.

Technology Leverage

Commitment to Innovation: Continuous emerging technology adoption and creation enables new product development and enhances product performance

Commitment to Creativity: Company leverages technology advancements to push the limits of form and function in the pursuit of white space innovation

Stage Gate Efficiency: Technology adoption enhances the stage gate process for launching new products and solutions

Commercialization Success: Company displays a proven track record of taking new technologies to market with a high success rate

Application Diversity: Company develops and/or integrates technology that serves multiple applications and multiple environments

Business Impact

Financial Performance: Strong overall financial performance is achieved in terms of revenues, revenue growth, operating margin, and other key financial metrics

Customer Acquisition: Customer-facing processes support efficient and consistent new customer acquisition while enhancing customer retention

Operational Efficiency: Company staff performs assigned tasks productively, quickly, and to a high-quality standard

Growth Potential: Growth is fostered by a strong customer focus that strengthens the brand and reinforces customer loyalty

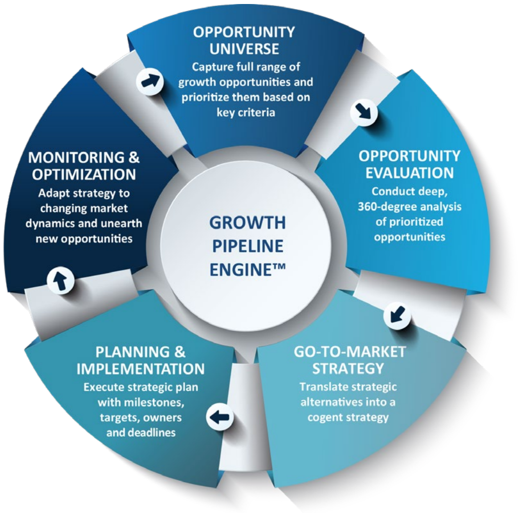
Human Capital: Commitment to quality and to customers characterize the company culture, which in turn enhances employee morale and retention

About Frost & Sullivan

Frost & Sullivan is the Growth Pipeline Company™. We power our clients to a future shaped by growth. Our Growth Pipeline as a Service™ provides the CEO and the CEO's growth team with a continuous and rigorous platform of growth opportunities, ensuring long-term success. To achieve positive outcomes, our team leverages over 60 years of experience, coaching organizations of all types and sizes across 6 continents with our proven best practices. To power your Growth Pipeline future, visit Frost & Sullivan at http://www.frost.com.

The Growth Pipeline Engine™

Frost & Sullivan's proprietary model to systematically create ongoing growth opportunities and strategies for our clients is fuelled by the Innovation Generator™.
Learn more.

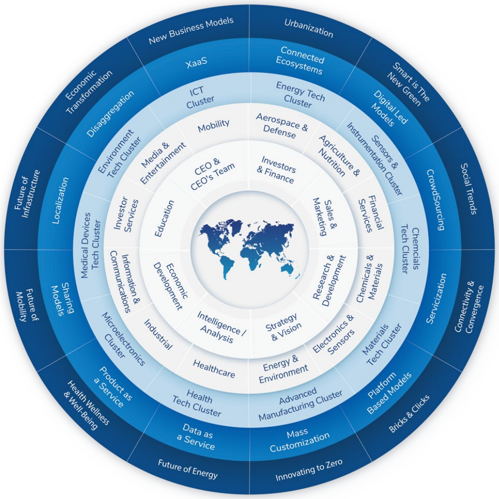

OPPORTUNITY UNIVERSE
Capture full range of growth opportunities and prioritize them based on key criteria
OPPORTUNITY EVALUATION
Conduct deep, 360-degree analysis of prioritized opportunities
GO-TO-MARKET STRATEGY
Translate strategic alternatives into a cogent strategy
PLANNING & IMPLEMENTATION
Execute strategic plan with milestones, targets, owners and deadlines
MONITORING & OPTIMIZATION
Adapt strategy to changing market dynamics and unearth new opportunities
GROWTH PIPELINE ENGINE™

Key Impacts:

- **Growth Pipeline:** Continuous Flow of Growth Opportunities
- **Growth Strategies:** Proven Best Practices
- **Innovation Culture:** Optimized Customer Experience
- **ROI & Margin:** Implementation Excellence
- **Transformational Growth:** Industry Leadership

The Innovation Generator™

Our 6 analytical perspectives are crucial in capturing the broadest range of innovative growth opportunities, most of which occur at the points of these perspectives.



Analytical Perspectives:

- Mega Trend (MT)
- Business Model (BM)
- Technology (TE)
- Industries (IN)
- Customer (CU)
- Geographies (GE)