Quiet Power at the Edge of Medicine

How the Kontron CG2500 Redefines Medical-Grade Edge Computing

White paper

kontron
The Power of IoT

Table of Contents

Executive Summary

Modern sequencing and medical imaging workloads require high-throughput, low-latency computation directly at the point of data generation. Centralized and cloud-dependent architectures introduce delays and performance bottlenecks that slow diagnostic pipelines, increase noise and heat in controlled environments, limit system scalability, and drive higher data egress costs. Kontron's CG2500 addresses these challenges with a compact 20-inch edge platform engineered for demanding medical applications, offering flexible CPU and GPU configurations to scale AI performance as workloads evolve.

Benchmark data shows the CG2500 delivers up to 3⁄ higher processing throughput, supports 80–111 active CPU threads, and fully powered GPU configurations—without the thermal throttling, forced core reductions, or acoustic penalties seen in competitive systems. Operating at 65–66 dB, it is also the quietest edge server in its class, enabling placement directly in sequencing labs and imaging suites where noise and reliability matter.

This white paper examines how medical OEMs, imaging specialists, and genomic researchers can leverage the CG2500 to increase throughput, improve data accuracy, reduce latency, and support next-generation instruments with greater efficiency at the medical edge.

Introduction — Why Noise, Power, and Reliability Matter in Clinical Environments

Clinical and research environments are changing rapidly. High-throughput sequencing instruments, advanced imaging systems, and AI-assisted diagnostics now generate terabytes of data per run—far beyond what traditional server rooms or cloud- dependent architectures were originally designed to support. As workloads grew, so did the need for systems capable of processing data directly at the point of acquisition, without latency or workflow disruption.

Edge computing emerged in response to this shift, allowing laboratories and imaging suites to move computation out of isolated server rooms and closer to where data is created. But this also introduced new operational constraints: servers must be compact, quiet, power-efficient, and reliable enough for continuous use in clinical spaces where uptime, temperature stability, and acoustic performance directly influence staff performance and patient experience.

The Kontron CG2500 was developed for exactly these conditions. Its high-performance, multi-GPU architecture and 8-10+ year lifecycle stability, combined with a 20-inch short-depth chassis, make it ideal for medical environments where space is constrained and reliability is essential. Benchmarking confirms it is the quietest and most powerful compact medical edge server available today capable of sustained throughput without thermal throttling or forced CPU reductions.

For medical OEMs, imaging manufacturers, and genomic research teams, this means the ability to:

- Manage massive data volumes at the edge
- Improve processing accuracy and speed
- Reduce latency and bandwidth dependence
- Maintain control of sensitive patient data
- Deploy high-performance compute directly within clinical workflows

Engineering Innovations: How the CG2500 Achieves Quiet Power

Delivering high-performance computation in medical environments requires a system-level approach to acoustics and thermal management. In the CG2500, quiet operation is achieved through the combined design of optimized fan geometry, engineered airflow ducts, material selection, and software-controlled cooling behavior. The ducts create defined cooling zones that work in coordination with dynamically controlled fans to move heat efficiently without excessive airflow or noise. Optimized heat-sink design further supports sustained CPU and GPU performance while maintaining low acoustic output suitable for clinical and laboratory environments.

1. Optimized Airflow Architecture for Low Acoustics

Traditional high-density servers rely on high-RPM fans that generate disruptive noise levels, making them unsuitable for sequencing labs, radiology workstations, or clinical environments. The CG2500's acoustic performance comes from:

- A short-depth 20-inch chassis with enhanced airflow management
- High-efficiency, hot-swap fans tuned for lower RPM operation
- Intelligent airflow channels targeting CPU, GPU, and PCIe components
- Air pressure design that maintains cooling efficiency without excessive fan noise

This enables operation at 65–66 dB, compared to ~72 dB in competing systems, creating a quieter clinical environment.

2. Power Delivery for Fully Utilized GPUs and CPUs

Many competitive edge systems require disabling CPU cores or limiting GPU power profiles to maintain system stability. The CG2500 avoids these compromises through:

- 2500W 80+ Titanium AC or DC redundant power supplies
- Support for fully powered NVIDIA L40S, L4, and A300 GPUs
- Optimized power delivery across multi-accelerator configurations
- Engineered to use less total power consumption
- Configurable CPU core allocation to match workload requirements and reduce power draw

3. Thermal Stability for Continuous Clinical Operation

The CG2500 maintains stable performance through optimized heatsinks, intelligent fan control algorithms, and balanced thermal distribution across up to 7 PCIe slots. This ensures zero thermal throttling during sustained benchmarking.

4. High-Efficiency Architecture in a Compact Footprint

The CG2500 avoids the noise and thermal penalties of dual-socket servers with a single-socket architecture supporting up to 64 or 128 cores. Its 20" compact size fits into sequencing carts, imaging equipment racks, and constrained clinical rooms.

5. Noise-Controlled Performance for Clinical Deployment

Noise influences equipment placement, staff fatigue, and patient comfort. The CG2500's engineered acoustic profile enables deployment directly beside sequencing instruments or radiology workstations without disruption, using»large-diameter fan blades and mechanical designs that efficiently distribute airflow throughout the chassis.

Performance Validation

Technical Benchmark Findings

Medical Imaging Benchmark: competitive overview

	Kontron CG2500	Top Leading Server
CPU	Single-socket 64 cores / 128 threads	Dual-socket 2x32 cores (only 48 active)
GPU Configurations Tested	L40S + L4 / L4 only / L40S + L4 + A300	L40S + L4
Performance (Gb/hr)	6 ▸ 12 ▸ 17 ▸ 23 Gb/hr (depending on configuration)	8–9 Gb/hr
Performance Gain	Up to 3x faster	Baseline
CPU Bottleneck	Resolved (80–111 threads active)	Significant (CPU-limited)
Noise Level	65–68 dB	72 dB
Power Handling	Fully powered GPUs (no limitation)	Limited – forced to disable 16 cores
Thermal Stability	Optimized airflow, lower dB	Higher thermal and acoustic footprint

Based on 2025 testing vs our competition, Kontron CG2500 outperforms in key metrics.

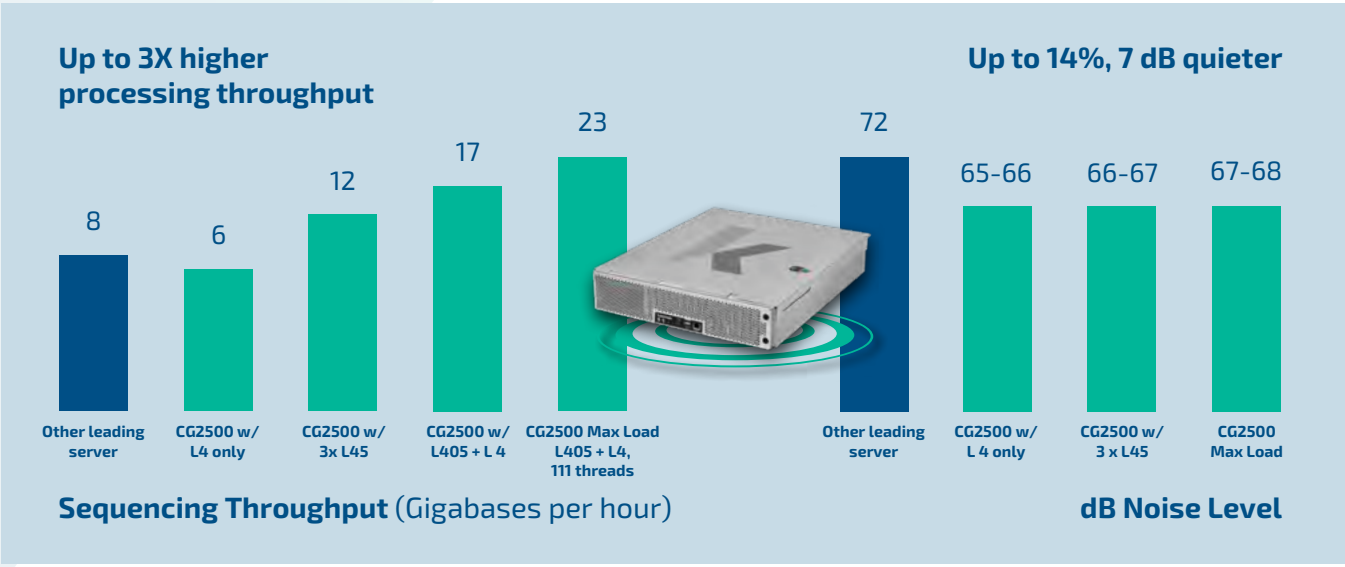
Benchmarking in 2025 compared the Kontron CG2500 with another leading edge server widely used in medical environments. Results showed consistent advantages in sequencing throughput, acoustic performance, and sustained compute stability.

Testing performed by a Kontron customer showed that the CG2500 can achieve up to 23 Gb/hr depending on CPU and GPU configuration—up to three times the throughput of the comparison system. This improvement is tied to more efficient CPU utilization (up to 111 active threads) and the ability to operate GPUs at full power without limiting cores or reducing performance.

These findings were particularly relevant for workloads such as next-generation DNA sequencing and high-volume medical imaging, where total processing time and system responsiveness directly impact laboratory throughput.

Acoustic measurements also showed that the CG2500 operates between 65–68 dB, making it 5–7 dB quieter than the other server tested.

In medical labs where noise can influence equipment placement and working conditions, this reduction offers a meaningful operational benefit.



Thermal assessments indicated that the CG2500 maintained stable temperatures under load, while the comparison system exhibited more thermal and power constraints, including restrictions on active CPU cores.

The tests also highlighted the importance of flexible GPU configurations in medical processing workflows. The CG2500 supported multiple validated combinations including NVIDIA L4, L40S, and A300, allowing performance levels to be matched to the specific needs of sequencing, imaging, or hybrid compute tasks.

Overall, the findings indicate that architecture, power management, and system design choices contribute significantly to measurable performance differences in edge servers used for medical applications.

Certifications & Compliance

The CG2500 meets major global compliance and safety requirements, including:

USA/Canada
UL 60950-1, CSA C22.2

Europe
CE Marking, EN 62368-1

International
IEC 62368-1 CB Report

Environmental
RoHS Compliant

Supported by ISO 13485-certified manufacturing for seamless integration into medical-grade devices.

Cybersecurity in Regulated Medical Environments

In regulated medical environments, cybersecurity is achieved through a combination of platform capabilities and system-level implementation. The CG2500 serves as a secure, server-class foundation by enabling local edge processing, isolated server management, and long lifecycle stability, helping medical OEMs eliminate data exposure and maintain controlled system behavior. Final security posture and regulatory alignment, including HIPAA considerations, are established through system integration, operating system hardening, and application-level security controls.

Built in North America — Reliable EMS Support

Manufactured in Kontron's Canadian facility, the CG2500 benefits from a stable North American supply chain, reduced tariff risk, faster turnaround times, and ISO-certified manufacturing processes. When required, Kontron can also support assembly in the United States or other global locations, providing medical OEMs with flexible, compliant manufacturing and deployment options worldwide.

Use Cases:
Genomics and Imaging Applications

Flexible and scalable: Different types of DNA sequencing methodologies.

Use Case #1
High-Throughput DNA Sequencing

Problem

Genomic research labs must process enormous volumes of sequencing data—often terabytes per project—to support applications such as precision medicine, rare-disease analysis, and population genomics. Traditional compute workflows depend on off-site servers or cloud resources, creating bottlenecks in data transfer, increasing processing times, and raising concerns around data integrity, security, and compliance. These delays slow down critical research outcomes. In addition, sequencing equipment generates data continuously, requiring compute platforms that can operate reliably in a laboratory despite temperature fluctuations, vibration, and strict noise constraints.

The same box could achieve long sequences with high precision and reduce the quickest sequence measurement with the same precision. It's adaptability to the exact configuration of the box with the customers. The CG2500 is flexible and scalable.

Solution

Deploying the CG2500 at the edge, directly within the research facility, allows labs to perform high-speed analysis of raw sequencing output on-site. This reduces full sequencing-to-analysis cycles from days or weeks to hours. The platform's high compute density supports AI (artificial intelligence)/ML (machine learning) workloads for tasks like base-calling, pattern detection, and variant identification, resulting in faster and more accurate insights. Its short-depth, low-noise design enables secure placement close to sequencing instruments, ensuring rapid, high-bandwidth data ingestion without relying on external networks. Modular accelerator cards can be added or swapped as sequencing pipelines evolve, giving labs the ability to scale performance without reworking their infrastructure. The CG2500's ruggedized architecture ensures continuous, deterministic operation—critical in medical environments where downtime is unacceptable.

Use Case #2
Medical Imaging Centers and AI Workflows

Problem

Diagnostic imaging facilities are generating increasingly large CT, MRI, and 3D imaging datasets as they adopt AI-assisted radiology tools. These workloads require high-throughput processing, low-latency inference, and strict protection of patient information. Relying on cloud-based analysis introduces transfer delays, inconsistent latency, and potential compliance risks—especially when handling PHI governed by HIPAA or regional privacy regulations. Imaging centers also need compute solutions that can operate quietly and reliably within clinical spaces, without disrupting patient care or requiring extensive IT infrastructure.

Solution

Placing the Kontron CG2500 at each imaging site creates a local, high-performance compute backbone for AI-driven radiology workflows. Large imaging files can be processed and analyzed directly on-site, reducing diagnostic turnaround times by removing the dependency on remote servers. With support for GPUs and dedicated AI accelerators, the platform enables real-time or near-real-time inference for anomaly detection, triage workflows, and image enhancement models. Its modular design allows imaging centers to integrate next-generation accelerator cards or higher-bandwidth storage as modalities advance, ensuring long-term scalability. The CG2500's short-depth architecture and low acoustic output make it suitable for installation near radiology workstations or imaging suites.

By keeping computation at the edge, facilities minimize bandwidth usage, maintain tighter control of sensitive patient data, and give radiologists dependable performance to support faster, more accurate decision-making.

Medical Imaging Benchmark: competitive overview

Capability	Kontron CG2500	Leading Competitor
Max Throughput	Up to 23 Gb/hr	8–9 Gb/hr
Performance Gain	Up to 3x	Baseline
CPU Utilization	80–111 threads	Bottlenecked
Required Core Disabling	None	Up to 16 cores
Noise Level	65–66 dB	~72 dB
GPU Support	L4, L40S, A300	Limited
Thermal Stability	No Throttling	Thermal Constraints

Conclusion — Setting a New Standard for Medical Edge Performance

As sequencing, imaging, and AI-driven diagnostics continue to advance, the need for reliable, quiet, high-performance edge computing becomes increasingly essential. The Kontron CG2500 delivers a rare combination of throughput, acoustic performance, thermal stability, scalability, and long lifecycle availability—all in a compact form factor suited for clinical environments.

With up to 3x performance gains, industry-leading low noise levels, and fully powered GPU configurations, the CG2500 enables medical OEMs, imaging centers, and genomic laboratories to process more data, more accurately, and more efficiently—directly where it's needed.

Summary of Key Design and Deployment Advantages

Performance and Efficiency

- High-performance compute for AI-driven genomics and imaging
- Flexible CPU and GPU configurations
- Low power consumption through efficient resource utilization

Reliability and Lifecycle Stability

- Rugged, server-class design for continuous operation
- Extended 8–10-year lifecycle support and BOM stability

Acoustic and Physical Design

- Low-decibel operation for clinical environments
- Compact 2RU short-depth rackmount design

Manufacturing, Security, and Compliance

- ISO 13485–certified manufacturing support
- Built and designed in North America (Canada) with U.S. EMS capabilities
- ITAR-free platform with cybersecurity-enabled system design

Experience the Quiet Power of the CG2500

Contact Us
info.americas@kontron.com

Kontron Canada Inc.
4555, rue Ambroise-Lafortune
Boisbriand (Québec) J7E 4K6

Kontron America Inc.
9477 Waples Street
San Diego CA 92121
USA

kontron
The Power of IoT

kontron.com