

Greater PCIe 4.0 Connectivity Drives Flexibility and Performance for AI at the Edge

Intel® Xeon® Scalable processor-enabled server boards deliver a powerful offering for solution providers in healthcare markets and beyond



"We chose 3rd Gen Intel® Xeon® Scalable processors because of their advanced features and higher performance. More performance means more expansion, more add-in cards, more flexibility."

—Ken Lin, senior product manager, DFI

Challenge: AI complexity and long service life

Artificial intelligence (AI) capabilities are widely recognized as a tool to help streamline radiology workflows—and reduce the radiologist's workload.¹ Many hospitals have already deployed AI-enabled solutions to automate medical image segmentation and labeling, but these solutions often rely on edge servers with expensive GPUs or other discrete accelerator cards. These devices must also undergo lengthy certification processes prior to deployment and are expected to stay in service for an extended period of time. In order to take full advantage of AI capabilities at the edge, healthcare IT providers need a flexible solution that is cost efficient and offers long-life availability.

Solution: IoT server boards with AI acceleration built in

With these requirements in mind, DFI, a global provider of embedded computing solutions, developed the ICX610-C621A server board enabled by 3rd Gen Intel® Xeon® Scalable processors. This flexible platform allows system integrators to design custom, powerful, and more-secure AI-enabled solutions for medical environments while supporting other use cases like industrial manufacturing. With integrated AI acceleration technologies like Intel® Deep Learning Boost (Intel® DL Boost) on the CPU, customers can deploy edge servers with the ICX610-C621A board as a cost-effective advancement over GPU-heavy solutions and still drive valuable and uniform medical-scan labeling to streamline radiologist workflows. "We chose 3rd Gen Intel Xeon Scalable processors because of their advanced features and higher performance. More performance means more expansion, more add-in cards, more flexibility," says Ken Lin, senior product manager at DFI. The ICX610-C621A server board supports Intel Xeon Scalable processor SKUs with long-life availability,² so hospitals can rely on product availability for years after their initial deployment.

Fast AI inference on images in medical and industrial applications

The 3rd Gen Intel Xeon processors deliver up to 1.56x improvement in AI inference for image classification compared to the previous generation.³ Lin also comments that DFI has been building AI systems using Intel® architecture for years, stating, "Most of our systems and motherboards are already powered by Intel® CPUs. This allowed us to leverage existing CPU-based infrastructure to run AI inference workloads." Embedded

technologies such as Intel® Advanced Vector Extensions 512 (Intel® AVX-512) help increase the performance of inference workloads on Intel architecture.

Intel DL Boost leverages Vector Neural Network Instructions (VNNI), an extension of Intel AVX-512 that improves performance by combining three instructions

into one. This offers a dramatic speed increase for targeted workloads, including image recognition, image segmentation, and object detection that are common to AI-assisted medical image analysis.⁴ It can also apply to any industrial use case that involves analyzing visual data.

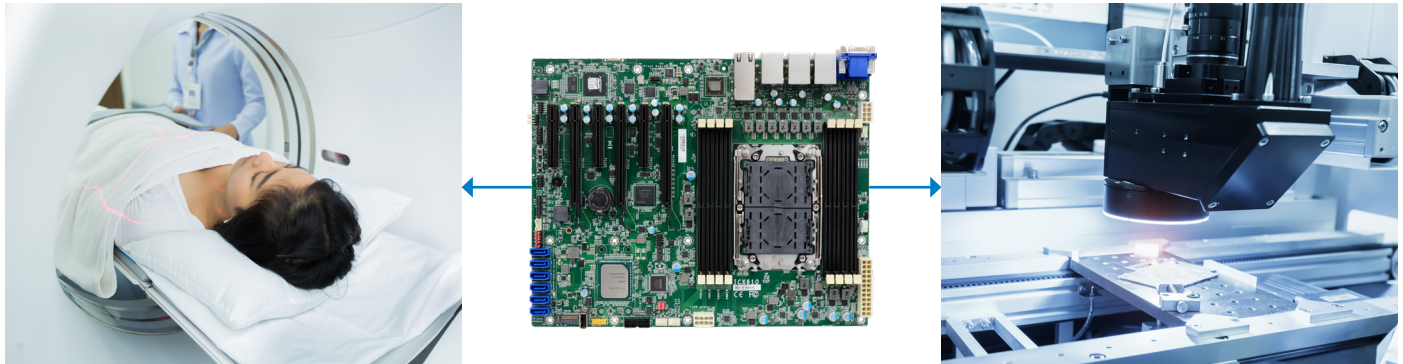


Figure 1. The ICX610-C621A is a foundational board to support medical, industrial, or other use cases for deep learning.

How it works

The ICX610-C621A server board with 3rd Gen Intel Xeon Scalable processors is built for edge server deployments. In hospital environments, the ICX610-C621A can be deployed in appliances at the IoT edge, integrated with medical scanners and equipment like MRI/CT machines, or implemented in nearby edge servers. With this offering, DFI customers get access to an Intel-enabled line of server boards with robust, power-efficient performance, high memory capacity, and PCIe 4.0 connectivity. System integrators (SIs) and solution providers can use this board with a broad selection of add-in cards to customize their portfolio while supporting deep learning and visual inspection use cases. DFI also offers guidance and expertise to help system builders configure their APIs and operating systems for the ICX610-C621A board and can optimize drivers, firmware, and BIOS settings for seamless functionality.

Why Intel Xeon Scalable processors?

Equipment manufacturers are looking for a single high-performing system that can support a wide variety of use cases. The 3rd Gen Intel Xeon Scalable platform does exactly that. The platform provides a high core count⁵ and 1.46x average performance gains vs. the prior generation,⁶ with an expansive set of technology features that enable a variety of use cases. These improvements are essential to DFI, which provides embedded computing hardware to clients in several industries all over the world. But the most important feature in the 3rd Gen Intel Xeon Scalable

processor family is support for more PCIe 4.0 lanes. "This platform supports up to 64 lanes of PCIe 4.0 per socket, which allowed us to build three PCIe 4.0 16-pin slots into the ICX610-C621A server board," Lin states. (The server board also supports two 8-pin PCIe 4.0 slots.) This flexibility allows system architects to customize their solutions with the ideal combination of add-in cards.

For example, if an end customer needs additional AI inference performance at the edge, a system integrator can use the ICX610-C621A server board to plug in multiple GPUs with a direct PCIe 4.0 connection to the processor.

More PCIe 4.0 lanes are an important factor to solution providers, in part because the majority of the deployments that DFI sees use a minimum of 3x 16-pin slots. In the 3rd Gen Intel Xeon Scalable platform, these are dedicated PCIe lanes for I/O operations. This platform uses Intel® Ultra Path Interconnect (Intel® UPI) for socket-to-socket communication, so platform controller hub (PCH) signal traffic doesn't use up valuable PCIe lanes. "The common use case is two capture cards to capture the image, with one GPU to support AI inference on 4K image scans," Lin says. "If we don't have the minimum of three slots, then solution builders would need to pick a dual-processor board." As a result, system builders who choose the ICX610-C621A can benefit from reduced hardware costs with a simpler single-socket configuration compared to a dual-socket board.

Security all the way down to the silicon

Data security is critical for hospitals in protecting patient confidentiality. 3rd Gen Intel Xeon Scalable processors provide hardware-enabled security features that help create a foundation of trust in the silicon. Intel® Total Memory Encryption (Intel® TME) enables full memory encryption, and Intel® Software Guard Extensions (Intel® SGX) isolate data in memory to help defend against malware and privilege-escalation cyberattacks.

Long-life availability

Healthcare institutions may keep appliances and equipment in service for several years beyond typical technology refresh cycles. Long-life availability is important to end customers for ongoing support, service, and replacements; to increase the value from equipment certifications; and to integrate new deployments seamlessly. Select SKUs in the 3rd Gen Intel Xeon Scalable processor family come with long-life availability,² specifically for IoT use cases where longevity is a top priority.

Recognition and stability with Intel partnership

The global market for AI in healthcare solutions is expected to grow to USD 31.3 billion by 2025, at a CAGR of 41.5 percent.⁷ Although use cases involving image segmentation and object recognition can help improve clinician efficiency, technological complexity and the high cost of performant, efficient IoT edge systems for AI are a high barrier to entry for many healthcare providers. Embedded solutions powered by DFI and Intel can provide a path for healthcare system integrators to enable new capabilities for their customers.

Throughout their decades-long partnership with Intel, DFI has taken advantage of Intel's Early Access Program to help align their portfolio of offerings with new product announcements. "Ninety percent of our customers already know Intel platforms," Lin says. "Sometimes they come to us for dedicated Intel products and the latest Intel offering. That's really helped us to do our business." DFI also benefits from Intel guidance and expertise with training for new technologies and quick answers via Intel® Premier Support. This has led to more-stable, more-reliable products overall. Lin says, "I can see our partnership growing as we work on more designs together."

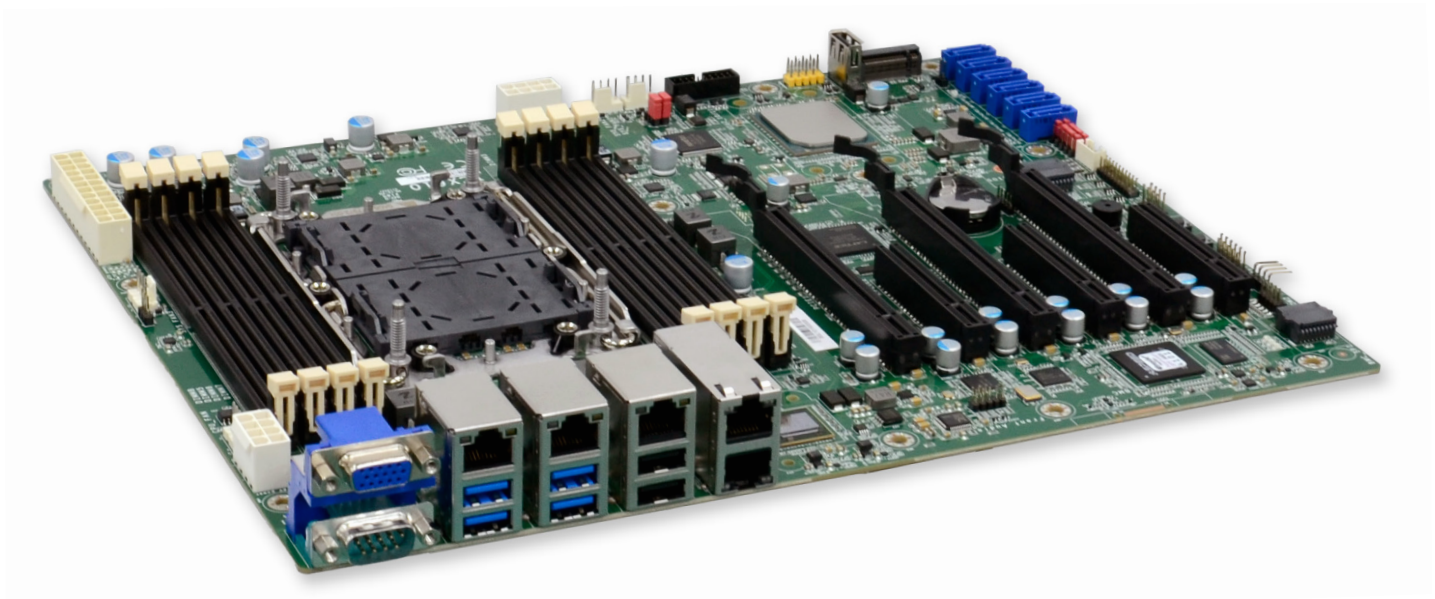


Figure 2. The ICX610-C621A offers 3x 16-pin PCIe 4.0 slots and 2x 8-pin PCIe 4.0 slots for greater customization.

Learn more

ICX610-C621A from DFI

This edge server board for AI IoT deployments allows for greater customization with support for more accelerator cards and higher CPU performance⁵ from 3rd Gen Intel Xeon Scalable processors.

[Learn more ›](#)

3rd Gen Intel Xeon Scalable processors

With IoT SKUs available, 3rd Gen Intel Xeon Scalable processors deliver higher performance⁵ and memory bandwidth and more PCIe 4.0 lanes for more-customizable deployments.

[Learn more ›](#)

About DFI

Founded in 1981 and based in Taiwan, DFI is a global provider of embedded computing solutions that focus on exceptional reliability and long-term life cycles across multiple industries.

[dfi.com](https://www.dfi.com)



1. See rsna.org/news/2019/july/ai-applications-address-workload.
2. Intel may change availability of products and support at any time without notice. Please contact your Intel account rep for additional information.
3. See [121] at intel.com/3gen-xeon-config. Results may vary.
4. See slides 5–8 at intel.com/content/dam/www/public/us/en/documents/product-overviews/dl-boost-product-overview.pdf.
5. The 3rd Gen Intel® Xeon® Scalable platform offers a maximum of 40 cores/socket; a maximum of 28 cores/socket are offered on the IOTG road map.
6. See [125] at intel.com/3gen-xeon-config. Results may vary.
7. "Artificial Intelligence In Healthcare Market Size, Share & Trends Analysis Report," Grand View Research, October 2019.

Notices and disclaimers

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Intel® Advanced Vector Extensions (Intel® AVX) provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause, a) some parts to operate at less than the rated frequency and, b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration, and you can learn more at intel.com/go/turbo.

Intel® processors of the same SKU may vary in frequency or power as a result of natural variability in the production process.

Performance varies by use, configuration, and other factors. Learn more at intel.com/performanceindex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.

Your costs and results may vary.

Intel® technologies may require enabled hardware, software, or service activation.

Not all features are available on all SKUs.

Not all features are supported in every operating system.

All product plans and road maps are subject to change without notice.

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

Intel is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's [Global Human Rights Principles](#). Intel's products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

0521/BC/CMD/PDF