Industrial manufacturers are conservative by nature. The value of their equipment and potential liability mean they generally won’t adopt new technology without it being a proven use case first.

So, while AI and the IoT will revolutionize IoT factory automation, many enabling technologies needed for industrial digital transformation must still make it from the drawing board into working proofs of concept (PoCs). This starts at the foundation of AIoT system architectures, where ideas like workload consolidation have yet to be successfully demonstrated at scale in real-world factories.

In fact, many automation professionals might still be unfamiliar with workload consolidation and why it’s important for achieving their smart-factory objectives. Simply stated, the term describes a way of virtualizing multicore processors. As a result, manufacturers can eliminate entire redundant systems, reduce total energy consumption, minimize latency, and lower costs.

This may seem simple, but it’s not. One reason workload consolidation hasn’t been sufficiently proven in automation use cases is that graphics processors leveraged in AIoT workloads like image processing and deep learning aren’t easily virtualized. Really, they can’t be virtualized at all without advanced features like the interface virtualization and I/O sharing designed into 12th gen Intel® Core™ processors (formerly “Alder Lake”).

These technologies are already being demonstrated in real-world industrial PoCs. But to fully appreciate them we must understand why they are required in the first place.

Virtual Graphics Are on the Outside Looking In

Although IoT virtualization isn’t new, some workloads are still easier to virtualize than others. Developers have struggled when attempting to virtualize graphics hardware because GPUs are usually host processor peripherals. And as specialized peripherals, GPUs can be exported from a host processor to another virtual machine (VM) but simply don’t have the features to natively support virtualization on their own.

In other words, GPU resources aren’t easily shareable across multiple VMs.

A developer could go to extreme lengths by emulating a virtual GPU that acts as an intermediary between the drivers and a physical GPU, but techniques like this add so much latency that most edge applications won’t tolerate it.

Helping Virtual GPUs Realize Their Full Potential

For workload consolidation and AIoT technologies to reach critical mass in the industrial sector, hardware-accelerated GPU virtualization is required. Fortunately for IoT factory automation professionals, virtual GPU performance improvement is one of many enhancements in 12th gen Intel Core processors.

Instead of addressing the problem with exotic architectures or more graphics execution units than an edge system could possibly use, these processors tackle it in the I/O that connects GPU peripherals. They do so by adding support for a PCI-SIG specification called Single-Root I/O Virtualization (SR-IOV) to Intel® Graphics Virtualization Technology (Intel® GVT), which gives VMs access to the physical functions of a GPU’s PCI Express port.
This makes one GPU that can itself be distributed across a workload-consolidated system at near-native performance levels. Simply put, it implements resource sharing directly in hardware rather than entirely in software.

12th gen Intel Core processors are the first to support both the Intel® Xe GPU architecture and SR-IOV virtualization features. And the global supplier of high-performance computing technology DFI, Inc., provides the ADS310-R680E, a microATX board, which is the first platform to equip GFX SR-IOV functionality. It also supports up to four external displays, Intel® OpenVINO™ Toolkit deep learning, and the Linux Kernel-based Virtual Machine (KVM).

A Proof Point for Industrial Workload Consolidation

For industrial automation equipment to make a seamless transition to AIoT it must add capabilities without sacrificing determinism. Optimizations like SR-IOV make this possible by allowing engineers to capitalize on the flexibility of modern software technology while still delivering native hardware performance, whether programs are executed on a physical host or virtually.

The ADS310 was recently part of a joint SR-IOV PoC with Intel, which demonstrated how graphics virtualization would perform in an industrial automation technology stack. In it, OpenVINO AI algorithms run in an Ubuntu container and analyze camera images, which are passed to a local monitor over HDMI. The same data is also fed into two Windows 10 OSs, partitioned by a KVM hypervisor, and then relayed to remote displays via Wi-Fi and HDBaseT Ethernet.

Without SR-IOV installed, the two Windows 10 instances achieved a 28 fps frame rate. With SR-IOV the VM frame rate jumped to a 60 fps frame rate, a common target for smooth graphics rendering.

The efficiency, productivity, ease of use, and cost benefits of moving to workload consolidated system architectures are both obvious and well documented. And now, thanks to the integrated capabilities of 12th gen Intel Core processors, they are also proven in the real world.

Automation industry, prepare to be transformed.